早稲田大学審査学位論文

博士（スポーツ科学）

Genome-wide Analysis of the Effect of Exercise on Translational Regulation and Co-expressed Gene Network

翻訳調節と共発現遺伝子ネットワークに対する運動効果の網羅的解析

２０１６年１月

早稲田大学大学院　スポーツ科学研究科

佐古　博皓

SAKO, Hiroaki

研究指導教員：　鈴木　克彦　教授

CONTENTS

PREFACE

The effect of exercise is highly sophisticated. To elucidate the mechanisms, intensive researches, including physiological, biochemical, and molecular biological approaches, have been conducted. However, there are still missing pieces in the approaches in exercise/sports sciences, especially at genome-wide levels.

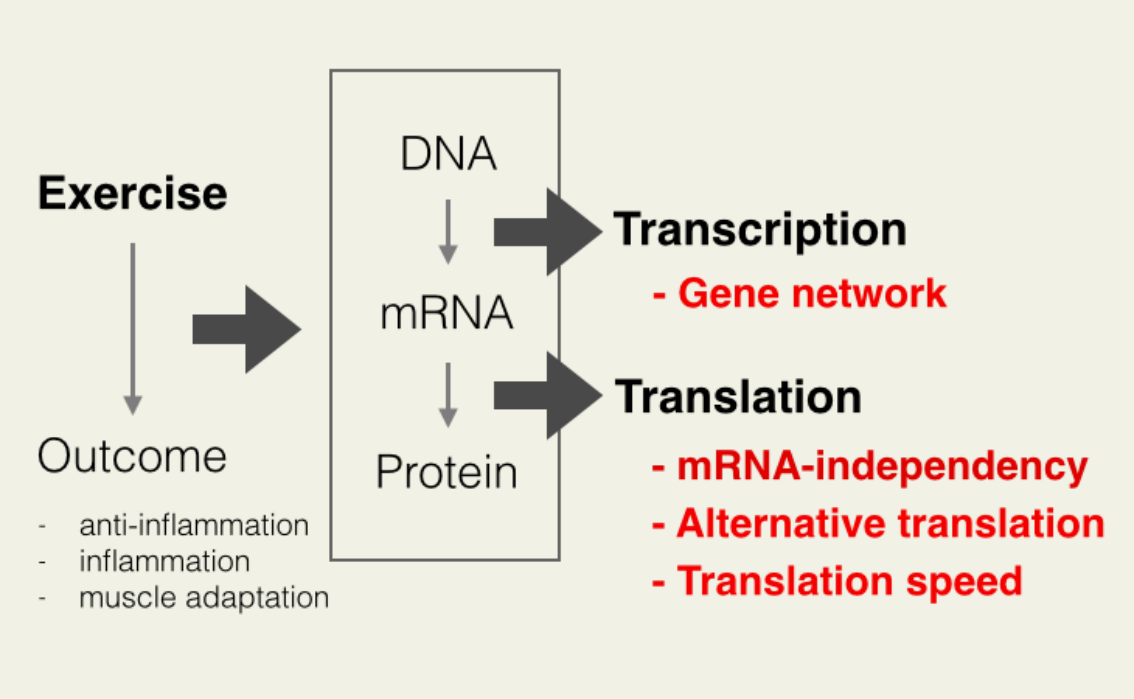
Genome-wide approach is a commonly used methodology to capture the global regulatory mechanisms. Although genome-wide gene expression or proteomics analyses have been carried out in exercise/sports sciences, there have been several limitations (Graphical abstract): 1. Translational regulation has been little focused and 2. Network-level dynamics of gene expression have never been investigated.

Translational regulation has been recently recognized as a critical determinant of the followings (Graphical abstract): 1-a. Protein abundance, 1-b. Protein species (alternative translation), and 1-c. Translation speed. Although they are hardly detected by

conventional analyses (e.g., gene expression or proteomics), they can exert pivotal influences on cellular homeostasis. Again, however, in the field of exercise/sports sciences, these genome-wide dynamics has never been investigated before.

Not only is it necessary to focus on genome-wide translational regulation, but also conventional genome-wide gene expression analysis could need to be re-considered. This is because conventional or widely recognized methods to analyze genome-wide gene expression lack in network dynamics of co-expressed genes (Graphical abstract). In nature, genes never work independently, but they do orchestrate together as a network to maintain cellular homeostasis. Despite the significance of the network dynamics, conventional gene expression analysis is unable to take into account of the concept.

Thus, in my thesis here, I aim to fill in the gaps in exercise/sports sciences particularly focusing on the following genome-wide analyses: 1. Translational regulation and 2. Network dynamics (Graphical abstract).



Graphical abstract

I. INTRODUCTION

I-i. Modification and Application of Ribosome Profiling to Investigate the Translational Regulation in Mouse Working Muscle

Pros and Cons in Conventional Genome-wide Analysis and Translational Regulation

Recent technological improvements have advanced analytical methods in all the fields of biology, including exercise/sports sciences. For example, proteomics and microarray have derived a lot of benefits to analyze protein abundance and gene/micro RNA (miRNA) expression in a genome-wide scale, respectively (*1*). The large-scale analysis enables us to capture molecular dynamics from a systematic perspective.

However, it is also true that both proteomics and microarray have their own drawbacks (*1*). As for proteomics, one limitation is the underestimation of low-abundance proteins (*2, 3*). In mass spectrometry-based proteomics, for example,

two-dimensional gel electrophoresis and/or liquid chromatography are often used for protein separation at a high-resolution level. On a two-dimensional gel, the spot of a lowly expressed protein could be masked by other highly expressed proteins (*2, 3*). Physiologically meaningful cytokines, including interleukin (IL)-6 and tumor necrosis factor (TNF)-α, are produced at very low concentrations. Concentrations of other highly abundant proteins, exemplified by serum albumin and α_2-macroglobulin, are much higher than those of IL-6 or TNF-α by up to 12 orders of magnitude (*4-6*). Even though the removal of such highly abundant proteins using an immunoaffinity column may improve the sensitivity to detect less abundant proteins, the removal efficiency is merely 80 to 90 %, indicating that there is still a large excess of 10^{11} the proteins (*4, 7*). Besides, proteins physically interacting with the targets to be removed, might also be filtered out, resulting in an unfavorable sample bias (*7*). Taken together, these issues may lower the sensitivity to identify and underestimate low-abundance proteins in a large-scale proteomics.

In a case of microarray, although such low-abundance gene expression can be

detected, post-transcriptional modification is ignored (*8*). Comparison between microarray and protein abundance has indicated low correlations between them (R^2 = 0.17 ~ 0.41), suggesting the insufficiency of mRNA expression as an indicator to estimate the corresponding protein abundance (*9-13*). Intriguingly, one study revealed that this low correlation is mainly caused by the ignorance of translational regulation (*14*). Consideration of translation remarkably improved the correlation from R^2 = 0.41 between mRNA and protein levels to R^2 = 0.95 between measured translation rates and protein levels (*14*). This finding indicates that translational regulation can exert significant effect on protein abundance and also that post-translation, such as protein degradation, has a weak effect (*14*). Thus, it is clearly evident that translational regulation is a critical determinant of protein abundance and biological phenomenon.

Analysis of Translation/Protein Synthesis in Exercise Science

In the field of exercise/sports sciences, the regulation of protein synthesis and amino acid metabolism is one of the major topics. Many researchers have intensively

investigated to understand the mechanisms of muscle adaptation to a variety of exercise modes and types, such as acute and chronic exercise as well as endurance and resistance exercise. Stable isotope tracers, for example, have enabled ones to examine the regulation of overall protein synthesis in skeletal muscle (*15-18*). During acute endurance exercise, overall muscular protein synthesis decreases in rodents and human (*19, 20*). The decrease in protein synthesis can be regulated by activation of AMP-activated protein kinase (AMPK) and decreased mammalian target of rapamycin (mTOR) signaling (*21, 22*). However, in spite of the intensive observation of overall muscular protein synthesis and widely recognized mTOR signaling cascades, little attention has been paid to a genome-wide skeletal muscle protein synthesis or translational regulation of individual genes. This is mainly attributable to the lack of methodology for a genome-wide analysis of translational regulation (*1*).

Ribosome profiling has recently made it possible to conduct a genome-wide translational regulation using a next generation sequencer (*8*). In ribosome profiling (*8, 23*), translating ribosomes can be stalled by translational inhibitor, followed by

enzymatic digestion of mRNA regions that are not protected by translating ribosomes. On the other hand, mRNA regions that are under translation (i.e., protected by translating ribosomes) cannot be digested. Therefore, these fragments are extracted and deep-sequenced by a next generation sequencer. The extracted fragments are called ribosome-protected fragments (RPF) with the length of approximately 30 nucleotides (nt) in mammals. After sequencing the RPF, sequenced RPF (or sequenced reads) can be mapped to a reference genome sequence to identify from which genes the RPF originated. As in a typical high-throughput mRNA sequencing (mRNA-Seq), in which highly expressed genes can be indicated by more mapped sequenced reads, genes with more mapped RPF suggest more active translation. Therefore, ribosome profiling makes it possible to conduct a genome-wide analysis of translational regulation focusing on individual genes rather than overall protein synthesis. I adapted ribosome profiling protocols to simpler *ex vivo* protocols using RAW264 macrophages (*1*), and I further applied the simplified ribosome profiling for *in vivo* mouse model to investigate the effect of a single bout of endurance exercise on translational regulation in mouse working muscle (in revision).

I-ii. Genome-wide Analysis of Translational Regulation: Alternative Translation

There can be multiple distinct translational isoforms generated from a single mRNA. Similarly to alternative splicing that results in multiple different mRNA isoforms generated from a single gene, alternative translation can synthesize protein isoforms or translational byproducts from a single mRNA. The isoforms include upstream open reading frame (uORF), N-terminal extension/truncation, and frameshift. uORF is an atypical ORF located in the upstream of the main ORF (mORF) and often much shorter than the mORF. Extension/truncation of N-terminal can be attributable to alternative translational initiation from the same main coding frame. Frameshift occurs outside of the main coding frame (+1 or -1) and often synthesizes totally different polypeptides compared with the mORF products.

Physiological Significance of Translational Byproducts

uORF

uORFs are recognized as an important player in maintaining homeostasis. uORFs are translated from canonical/non-canonical start codons. In mammals, at least one uORF can be observed in roughly 50% of transcripts and often short in length (< 30 codons) (*24*) (*25*). An uORF is also known as a translational regulator of the downstream mORF. Most uORFs have negative effect on the downstream mORFs, in which translating uORFs can reduce ribosomal re-initiation and therefore decrease the translation rate of the mORF (*26-28*). For example, translation of β-site amyloid precursor protein cleaving enzyme 1 (Bace1) mRNA, mediating the excessive accumulation of β amyloid seen in Alzheimer's disease, can be regulated by its uORFs (*29*). Under healthy conditions, the 4th uORF of 6 Bace1 uORFs is actively translated, decreasing the translation rate of Bace1 mORF. This consequently represses β amyloid accumulation (*29*). These previous studies support the physiological significance of uORFs.

N-terminal extension/truncation

Extension/truncation has direct impacts on the corresponding protein functions. Diverse extensions and truncations, accounting for as much as 14 % of all the translated products, have been discovered in mouse embryonic stem cells (*28*). N-terminal truncation is of particular interest due to a possible lack of the functional domain, leading to dysfunction of the protein. In a case of transcription factor, for example, alternative translational initiation of CCAAT/enhancer binding protein (Cebpa) produces an isoform shorter than the full length Cebpa protein by 12 kDa (*30*). It has been reported that the truncated form has DNA-binding domain whereas it lacks in the transactivation domain, resulting in prevention of the original function of Cebpa anti-mitotic activity (*30*). This exhibits a clear evidence for the significant effect of protein extension/truncation.

Frameshift

Frame-shifted translation generates proteins totally different from the main products synthesized from the mORF. Although these byproducts can be degraded to prevent them from aggregating and damaging the cells, such frame-shifted byproducts can

possess molecular function. One example is Ornithine decarboxylase (ODC) antizyme (*31*). ODC antizyme is a negative feedback regulator of ODC, responsible for polyamine synthesis (*32*). Increasing concentration of polyamine triggers a +1 frameshift, generating frame-shifted ODC antizyme which can promote the degradation of ODC to maintain polyamine abundance (*31*).

Ribosome profiling also enables ones to detect these alternative translations. Here, to develop a bioinformatics pipeline to screen all the reference genome sequence and to identify such isoforms and byproducts, RAW264 macrophages were first used and I established the pipeline to further analyze mouse skeletal muscle. Application of the pipeline to RAW264 macrophages and mouse skeletal muscle revealed previously unrecognized translational isoforms and byproducts.

I-iii. Genome-wide Analysis of Translational Regulation: Dynamic Change in

Translation Speed of Ribosomes

Significance of Translation Speed and the Difficulty of the Investigation

Genome-wide analysis of translational regulation can not only reveal translational abundance or byproducts but also address whether the speed of translating ribosomes of each gene is altered. Translation speed is now recognized as a critical determinant to regulate protein conformation and even its function by affecting a nascent polypeptide folding (*33, 34*). Therefore, it is of importance to understand how translating ribosomes are stalled and where such stall occurs and also crucial to investigate whether pathophysiological changes, such as inflammatory or exercise stimuli, can alter the speed of translation.

However, it has been difficult to examine at a high resolution the effect of pathophysiological changes on translation speed. In particular, in terms of where translational stalls occur, there have been some discrepancies between *in vitro* and *in vivo* analyses. *In vitro*, structural basis cryo-electron microscopy (cryo-EM) analysis has

revealed specific interaction at near-atomic-resolution between nascent amino acid residues and ribosomal structure, such as nucleotide bases of ribosomal RNA (rRNA) protruded out of the ribosomal exit tunnel wall (*35-38*). Despite the high resolution, the effect of physiological changes is difficult to be examined *in vitro*.

Ex vivo, on the other hand, the use of ribosome profiling is suggested to estimate translation speed. Ribosome profiling exhibits ribosome densities, where denser ribosomes in a mRNA region suggest that these ribosome resided longer at the region, indicating slower movement of the ribosomes (*28*). The use of ribosomes profiling has discovered what factors affect translation speed, including codon bias/mRNA folding energies (*39*), adaptation of secretory protein gene families to transfer RNA (tRNA) pools (*40*), and specific nascent residues (*28, 41, 42*). However, the results addressing where such stalls occur are often inconsistent to those in *in vitro* observation. Arginine (Arg) residues, for example, have been widely recognized *in vitro* to induce translational stall via electrostatic interaction between the positively charged Arg residues and negatively charged ribosomal exit tunnel (*43, 44*). *Ex vivo*, however,

despite the detection of Arg-induced translational stalls (*45*), it has been difficult to accurately identify the stall sites inside the ribosomal exit tunnel.

Novel Analytical Approach to Examine Translation

I speculated that the discrepancy between *in vitro* and *ex vivo*, especially where inside the ribosomal exit tunnel stalls occur, could be attributable to the bias in translation. Because ribosome profiling is so robust and sensitive that one can detect unique translational events in a single mRNA (*8, 28, 46*), gene diversity (e.g., different base composition and gene expression levels) might hamper a high-resolution analysis of translational stalls. For example, in a case that small portion of highly abundant mRNAs dominate the total sequenced reads, detected global signals of translation may reflect only the small fraction of the mRNAs. This indicates that lowly expressed mRNAs little contribute to the obtained signals of translation. The highly biased observation might prevent the potentially interesting global characteristics and make it difficult to detect where in ribosome translational stalls occur.

To minimize the inter- and intra-gene bias and to observe translational stall at a high resolution, I developed a novel bioinformatics analysis, named binary count analysis. The results showed the detection of translational stall at an unprecedented resolution *ex vivo* and also in *in vivo* mouse skeletal muscle. The current approach further revealed unexpected dynamics in translational stalls induced by acute inflammatory stimuli. Taken together, in addition to the conventional view that different conditions (e.g., inflammatory or exercise stimuli) can change the expression levels of mRNAs or proteins, now it can be implied that different pathophysiological conditions might cause change in translation speed and therefore conformational/functional changes of nascent proteins, consequently feeding back to the regulation of pathophysiological conditions.

I-iv. Genome-wide Gene Expression Analysis: Gene Co-expression Network Analysis

Big Data and Drawbacks in Conventional Gene Expression Analysis

Recently, it has been easier to conduct a genome-wide analysis. There has been an increasing number of publicly available “big data”, including genomics, transcriptomics, translatomics, and proteomics analyses. Using such freely accessible database, it is now possible to take a systematic approach to further investigate genome-wide dynamics of genes and their expression. Most commonly used analysis is to detect the differential expression of mRNA. A standard workflow of differential expression analysis can be: (i) Appropriate statistical model selection, (ii) Normalization of the expression values, and (iii) Detection of differentially expressed genes (DEG). Comparing more than two different conditions, ones can detect DEG, in which genes showing the largest differences often have important pathophysiological impacts and therefore they are likely to become further research targets.

However, despite the simple and widely recognized DEG approach as I used to identify differentially regulated translation, there are several limitations. Major drawbacks of detecting DEG can be a strict statistical threshold and the lack of network

and hub concepts (*47, 48*). Genes showing moderate expression changes but below a statistical threshold (e.g., 2-fold changes in microarray) are often excluded from further analysis. However, when these genes are enriched in a specific signaling pathway, the pathway could be affected by the accumulated effects of the small expression changes (*48*). In other words, the genes forming a network, where a specific biological pathway is preserved, can have a biological importance even if their expression changes do not exceed a statistically significant threshold.

Gene Co-expression Network Analysis

Gene co-expression network analysis, not restricted by a conventional statistical threshold, can infer novel gene co-expression network from gene expression data by clustering co-expression of genes (*48-57*). Using Weighted Gene Co-expression Network Analysis (WGCNA) (*47*) and big data, such as microarray, it is possible to construct *de novo* weighted-gene co-expression networks by calculating Pearson correlations of the expression patterns of each pair of genes, followed by grouping highly correlated genes together, defined as a network module. One module forms one

network and often specializes in a specific biological function (e.g., innate immune response). Taken together with the expression changes of the genes in a network depending on different conditions (e.g., control vs. treatment), one can investigate gene regulation at a network level rather than an individual gene level (*47*).

Another beneficial aspect of gene co-expression network is the notion of "hub gene". In a scale-free network, for example, network of global flight paths, there are some hub airports (e.g., New York, Tokyo, Paris). A hub airport is connected by many flight paths, each of which is bound to a different airport. Therefore, such hub airports have important roles in the flight network and the loss of a hub airport would be a huge matter. Similarly in a gene co-expression network, a constructed co-expression network can generate hub genes, which can have a strong influence on the network (*47, 48, 58-60*). Although the concept of hub genes lacks in a conventional DEG analysis, WGCNA is able to suggest functionally critical hub genes even if they are previously uncharacterized.

Two distinct concepts, co-expression network and hub genes, generated by WGCNA enables us to investigate network preservation in different conditions. This is suitable to examine the effect of different modes of exercise in different tissues and different time-courses. Because different modes of exercise derive highly diverse effects in many different tissues and in different time-points after exercise, the conventional DEG analysis has been difficult to understand the dynamics at a systematic level. Taking advantage of the notion of co-expression network and hub genes, here using publicly available microarray datasets, I investigated how different exercise modes affect the network dynamics in different tissues and different time-courses and also analyzed to what extent the affected networks can be preserved among different exercise modes, tissues, and time-courses.

II. MATERIALS AND METHODS

II-i. Modification and Application of Ribosome Profiling to Investigate the Translational Regulation in Mouse Working Muscle

Accession number

The sequenced and processed files from the current study of mouse skeletal muscle have been deposited in the National Centre for Biotechnology Information's Gene Expression Omnibus under the accession number GSE69699.

Animals and endurance exercise

Seven-week-old male C57BL/6J mice, purchased from Kiwa Laboratory Animals (Wakayama, Japan), were first acclimated to the lab environment for a week (temperature controlled at 23 °C with 12 hr light/12 hr dark cycle and fed with normal chow). They were then acclimated to a motorized treadmill (Natsume, Kyoto, Japan) for approximately 10 min. After another one week, mice were randomly assigned into five different groups: no exercise (n = 6), immediately after exercise (n = 6), post 1 hr of exercise (n = 6), post 2 hr of exercise (n = 6), and post 4 hr of exercise (n = 6). The

exercise group mice underwent treadmill running at 20 m/min for 1 hr. To accurately identify where ribosome is located on mRNA, a translational inhibitor, cycloheximide, was used to stall the movement of ribosomes. At the indicated time points, under isoflurane anesthesia, 5 μl of 50 mg/ml cycloheximide (Sigma, St. Louis, MO) in PBS was injected into right gastrocnemius of all the mice (aged 9 to 10 weeks) and the right gastrocnemius was immediately harvested, followed by brief wash in 100 μg/ml cycloheximide in PBS and flash-frozen in liquid nitrogen. The cycloheximide-treated samples were further processed for mRNA-Seq and ribosome profiling as described below. Left gastrocnemius was harvested for western blotting and real time qPCR analyses. All the animal experiments were approved by the Institutional Animal Care and Use Committee at Waseda University (approval number: 2014-A085) in accordance with the guidelines of Ministry of Education, Culture, Sports, Science and Technology of Japan for animal experimentation at research institutes.

Total RNA extraction for mRNA-Seq and ribosome profiling.

The right gastrocnemius was homogenized in 1.2 ml of ice-cold lysis buffer [20 mM

Tris HCl pH7.4, 150 mM NaCl, 5mM $MgCl_2$, 1 mM DTT, 100 µg/mL cycloheximide (Sigma), TURBO DNase I 25 U/mL (Ambion, Life Technologies, Carlsbad, CA), and 1% Triton X-100 (Sigma)], followed by centrifugation at 15,000 g for 10 min at 4 °C. Supernatant from the lysate was separated in two 600 µl aliquots; one for mRNA-Seq and the other for ribosome profiling. Three sets of 100 µl of the aliquot from 3 individual mice were combined together to create 1 biological replicate (i.e., 1 biological replicate was composed of 3 different mice both in mRNA-Seq and ribosome profiling).

RNA fragmentation for mRNA-Seq

Extracted total RNA was purified by miRNeasy Mini Kit (Qiagen, Hilden, Germany), followed by poly(A) RNA extraction by Dynabeads mRNA DIRECT Micro Kit (Ambion). For fragmentation, purified mRNA was mixed with 2x alkaline buffer (90mM $NaHCO_3$, 10mM Na_2CO_3, 2mM EDTA) and incubated at 80°C for 15 min. The reaction was stopped by adding ice-cold solution (48.5µl RNase-free water, 1.5µl GlycoBlue (Ambion), and 10µl 3M NaOAc), followed by 150µl isopropanol and precipitated by standard protocols.

Monosome extraction for ribosome profiling

To extract the mRNA regions that were protected by ribosomes (i.e., mRNA fragments under translation), total RNA lysate 300µl was mixed with 15µl RNase If (New England Biolabs, Beverly, MA) and incubated for 45 min at room temperature to digest the mRNA regions that were not protected by ribosome. Reaction was halted by 10µl SUPERase·In (Ambion). To extract monosome with ribosome-protected mRNA fragments (RPF), digested sample was loaded onto sucrose cushion buffer (34% sucrose, 20mM TrisHCl, 150mM NaCl, 5mM $MgCl_2$, 1mM DTT, 100µg/ml cycloheximide, and 20U/ml SUPERase·In) with 1:3 ratio in a polycarbonate centrifuge tube (Beckman Coulter, Palo Alto, CA), followed by centrifuge for 4hr at 400,000xg, 4°C in MLA130 rotor (Beckman Coulter). Pellet was purified by miRNeasy Mini Kit and then precipitated by standard isopropanol precipitation protocols.

Size selection of fragmented mRNA and RPF

Purified RNA samples, small RNA II Marker (Funakoshi, Tokyo, Japan), and

synthesized oligo (Both 3' ends were phosphorylated and both 5' ends were dephosphorylated. upper: 5'-AUGUACACGGAGUCGAGCUCAACCCGCAACGA, lower: 3'-AUGUACACGGAGUCGACCCGCAACGA) (Greiner, Tokyo, Japan) were each mixed with 2x Novex TBE-Urea sample buffer (Invitrogen, Life Technologies, Carlsbad, CA) and denatured at 80°C for 90 sec. Samples were then loaded on a denaturing 15% polyacrylamide TBE-urea gel (Invitrogen) and run for 65 min, followed by SYBR Gold (Invitrogen) staining for 5 min. The gel regions of 26~32 nucleotides (nt) were excised for RPF recovery. The gel regions of 25~45 nt for fragmented mRNA were excised. Size-selected gel was disrupted and incubated in 360μl RNase-free water at 70°C for 10 min. All liquid and gel were transferred into a Spin-X column (Corning, NY) and centrifuged for 2 min at full speed, followed by standard isopropanol precipitation protocols.

ribosomal RNA (rRNA) removal, 5′ phosphorylation, and 3′ dephosphorylation from RPF.

RiboMinus (Invitrogen, Life Technologies, Carlsbad, CA) was used to remove rRNA

from RPF samples. For both fragmented mRNA and rRNA-deleted RPF, 5′ phosphorylation and 3′ dephosphorylation were carried out. The 3′ ends were first dephosphorylated at 37°C for 60 min in reaction mixture [10 U/μl T4 polynucleotide kinase (New England Biolabs, Beverly, MA), T4 polynucleotide kinase buffer without ATP, and 20 U/μl SUPERase·In (Ambion)], immediately followed by 5′ phosphorylation for 30 min in the presence of 1 mM ATP (New England Biolabs) and then RNA was participated.

Sequence library preparation, sequencing, and alignment.

I prepared sequence library using Ion Total RNA-Seq Kit v2 (Ion Torrent, Life Technologies, Carlsbad, CA) according to the manufacturer's protocols except that reverse-transcribed cDNA was gel-purified to remove excess primer dimers. The region (approximately 63 nt) was excised and recovered by the same protocols as described above. Sequencing was carried out using Ion PGM Template OT2 200 Kit, Ion PGM sequencer, Ion PGM Sequencing 200 Kit v2, and Ion 318 Chip Kit v2, according to the manufacturer's instructions (Ion Torrent). Base-calling and alignment were conducted

using the built-in Ion torrent software (V4.0.1) and the reads shorter than 24 nt were removed. Sequenced reads were mapped and aligned to coding DNA sequences extended to 25 nt upstream of translational start site of each gene, retrieved from GRCm38 using Biomart MartView (http://www.biomart.org/biomart/martview) (*61*). As for reads with multiple hit, the best hit was aligned. All the figures and statistics below were created by the R software, if not otherwise specified.

Reproducibility and validity of sequencing.

To compare the reproducibility of two independent biological replicates, aligned read counts were normalized by total read count (RPM: Reads Per Million). For validation of ribosome profiling, triplet periodicity was confirmed. For each gene, 5′ end position of aligned reads was categorized into 0, -1, or +1 relative position to the main coding frame and the ratios of each frame were calculated.

To determine minimum threshold of read counts for robust statistical analysis, read counts of one replicate were divided by total read counts composed of both replicates to calculate the fraction of one replicate. In a case of perfect reproducibility, the fraction

would be 0.5. Based on average read counts, genes were binned and SD of the fraction was computed for each bin. The SD was predicted using binominal partitioning of the read counts. The minimum read count, 125 RPM, was selected because variation between biological replicates becomes stably larger than that predicted from counting statistics approximately from 125 RPM.

Sequenced data analysis.

Focusing on 1011 genes (> 125 RPM), principal component analysis was carried out. I plotted PC1 and PC3, well representing the differences between the profiles of transcription/translation and no exercise/exercise, respectively. To detect differentially regulated genes at transcriptional and translational levels, edgeR was used with TMM normalization (*62, 63*). For the differentially regulated genes, pathway enrichment analysis was conducted using InnateDB (*64*) with hypergeometric algorithm and Benjamini Hochberg *P*-value correction where only the pathways in KEGG (*65*) were focused. As for the known TOP motif genes, I focused only on highly expressed genes (> 125 RPM). Translational efficiency was obtained by dividing normalized RPF counts

by normalized mRNA-Seq counts. Kolmogorov-Smirnov test was used for statistics.

Real time qPCR and western blotting.

Total RNA was extracted using TRIzol (Invitrogen), followed by cDNA synthesis using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Life Technologies, Carlsbad, CA) according to manufacturer's instruction. Solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 25 (Slc25a25) gene expression was determined using SYBR Green (Applied Biosystems) with primers of Slc25a25, Matrix metalloproteinase-2 (Mmp2), Nuclear Factor (erythroid-deived 2)-like 2 (Nrf2), Atrogin1, Muscle RING-Finger Protein-1 (MuRF1), Peptidase mitochondrial processing alpha (Pmpca), YME1-like 1 ATPase (Yme1l1), and IMP2 inner mitochondrial membrane peptidase-like (Immp2l) was used. Protein was extracted using PARIS kit (Ambion) supplemented with protease inhibitor cocktails (Thermo Fisher Scientific, Life Technologies, Carlsbad, CA) according to manufacturer's instruction. For western blotting, 5 μg of total protein was loaded into 4-12 % Bis-Tris precast polyacrylamide gel (Thermo Fisher Scientific), followed by electrophoresis and

transferring to PVDM membrane using iBlot 2 system (Thermo Fisher Scientific). Anti-SLC25A25 (GeneTex, Irvine, CA) and anti-GAPDH (Cell Signaling, Danvers, MA) for internal control were used and ECL (GE Healthcare, Pittsburgh, PA) was used for detection. The experiments were repeated to confirm the reproducibility. As for statistics, one-way ANOVA with Tukey was used. Using the expression levels of real time qPCR and western blotting, Pearson correlations were calculated between Slc25a25 protein abundance and other gene expressions. The correlation patterns were visualized by Cytoscape (*66*). I then carried out multiple linear regression analysis using the expressions in the groups: no exercise, immediately after exercise, and post 1 hr of exercise, in which all the genes were initially included to validate the model and then the least significant gene was removed repeatedly at each validation (i.e., Slc25a25 mRNA, Atrogin1, MuRF1, and Yme1l1 were removed).

II-ii. Genome-wide Analysis of Translational Regulation: Alternative Translation

Accession number

The sequenced and processed data used in the current analysis of RAW264 macrophages have been deposited in the National Centre for Biotechnology Information's Gene Expression Omnibus under the accession number GSE60930. The following analysis was carried out using in-house R scripts. As for the data of mouse skeletal muscle, the same datasets described above were used.

Cell culture

RAW 264 cells (DS Pharma Biomedical, Osaka, Japan) were cultured at 3.0×10^5 cells/mL in media (DMEM, 2 mM Glutamine, 10% FBS, 100 units penicillin and 100 μg streptomycin/mL) 24 h prior to harvest. The confluency of cells was confirmed to never exceed 80%–90%. Two independent biological replicates for both ribosome profiling and mRNA-Seq were prepared. Sequence library preparation steps for both RAW264 and mouse skeletal muscle were carried out as described above. To induce acute inflammation in RAW264, 100 ng/ml of lipopolysaccharide (LPS) was added into

media for 30 min.

Detection of N-terminal extension/truncation and translational frameshift

To identify alternative translation, an in-house analytical pipeline was created. For a robust analysis, I focused only on highly translated transcripts, at least 1,000 RPK (aligned reads per kb in a transcript). The aligned read positions in a reference gene and the corresponding annotated CDS start/stop positions of the gene were combined into one data file. The aligned read positions were allocated into one of three coding frames, namely frame0, frame1, and frame2. Aligned read numbers and positions in each coding frame were plotted as well as the annotated CDS position. Visually inspecting the aligned read distributions and annotated CDS positions made it possible to identify translational isoforms, including N-terminal extensions/truncations, uORFs, and frameshifts.

Western blotting

To confirm the presence of RAC2 translational isoforms, western blotting was

conducted with anti-RAC2 antibody (Abcam, Cambridge, UK). Importantly, the epitope of the RAC2 antibody (Abcam) was located at the C-terminal of RAC2 protein. Because the newly identified RAC2 alternative translations were distinguished by the N-terminal truncation, the C-terminal-specific antibody is able to detect both isoforms. Western blotting was carried out as described above.

II-iii. Genome-wide Analysis of Translational Regulation: Dynamic Change in Translation Speed of Ribosomes

Validation of Binary Counts

After sequencing and aligning the sequenced reads to the reference genome, I obtained the 5′ end positions relative to the start site in each gene. To normalize inter- and intra-gene bias, I removed identical 5′ end positions in each gene and converted its read counts into “1” count. The processed read count was named as “binary” count

because each base position in a gene takes only “0” or “1” read counts after the binary conversion. To validate the binary approach, cumulated coefficient variations (CV) for the main coding frame of RPF counts were calculated using all the detected genes.

$$CV_{accum.} = \frac{\sum_{k=1}^{L} SD_k}{\sum_{k=1}^{L} Mean_k}$$

$CV_{accum.}$: accumulated coefficient variation

L: length of genes

k: position of genes

Kolmogorov-Smirnov test was used to compare CV before and after applying binary density approach.

TAC site and translational stall detection by metagene analysis

To examine the effect of single residue of interest on translation speed, I defined sequence windows for the genome-wide metagene analysis similarly used in a previous study (*41*). In the sequence windows, there is only one amino acid of interest encoded. Importantly, there is no additional residue of interest encoded in the 10 codons 5′ side

of the residue and similarly no extra residue of interest in the 25 codons 3′ side of the residue. For example, when the effect of Glu residue on translation speed was examined, I focused on the reference sequences encoding single Glu residue and there was no extra Glu codon in 10 codons (30 nt) 5′ side of the target Glu as well as no additional Glu codon in 25 codons (75 nt) 3′ side. The length, 10 codons, is roughly equal to the width of ribosomal complex (*67*). Given the length of exit tunnel (100 Å) and single residue (3.35 Å in length) (*43, 68, 69*), the 35 codons length (10 + 25 codons) indicates the approximate distance from the entry of the ribosomal complex to the end of the exit tunnel, which allows the presence of only single residue of interest (e.g., Glu) in the ribosome. This can avoid the possible effect of the repeated occurrence of multiple Glu residues. After finding such sequence windows genome-wide for Glu, relative position (nt) to the Glu codons was assigned to each 5′ end of the aligned reads. I counted up the number of the identical 5′ relative positions genome-wide for ribosome profiling and mRNA-Seq and then divided the summed RPF counts by the counts from mRNA-Seq to calculate and normalize the ribosome density at every single base position. If Glu slows translation speed at a specific site in ribosomal complex, the stalled ribosome is

more likely to be captured as RPF in sequence windows. When all the genome-wide sequence windows are piled up, the base position with higher likelihood of producing RPF stochastically results in a higher ribosomal density peak compared to those in other positions, which indicates translational stalling.

Differentially Regulated Translational Stall and Trypsin Assay

Similarly, to identify translational stalls induced by a codon, I conducted a metagene analysis for 61 codons and additionally analyzed dynamics of the stalls derived from inflammatory and exercise stimuli. Differentially regulated translational stalls were identified by Student's t-test ($P < 0.05$ and FDR < 0.1) and absolute Z score differences (> 1.5). Absolute Z score differences were calculated as following: Z score difference = $| Z_{\text{Stm codon}} - Z_{\text{Bsl codon}} |$ (Stm: inflammatory or exercise stimuli, Bsl: basal, codon: codon of interest). After trypsin assay, in which 2.5 μg of total protein was digested by 0, 10, 20, or 30 ng/μl of trypsin (Sigma) at room temperature for 30 min, western blot was carried out with anti-Calreticulin (Abcam) and anti-GAPDH (Cell Signaling) to confirm the effect of decreased translational stalls in Calreticulin with negative control GAPDH

in which much less translational dynamics was observed.

II-iv. Genome-wide Analysis: Gene Co-expression Network Analysis

Data Collection and Pre-processing

All the microarray raw data of human mRNA expression analyses was collected from the National Centre for Biotechnology Information's Gene Expression Omnibus. I first screened 40 microarray datasets. To minimize the bias induced by batch effect, I mainly focused on the data generated by Affymetrix platforms because most of the exercise studies have been carried out by Affymetrix platforms (21 datasets). I also removed the datasets that lacked in clear description of samples and narrowed down to 5 datasets that suited for the current research design (database IDs: GSE1140, GSE28422, GSE45426, GSE51837, and GSE59088). Two studies examined blood cells after acute endurance

exercise (immediate after and post 1hr). One study was for vastus lateralis (VL) after acute endurance exercise (post 2.5hr and 5hr). Another study was for VL after acute resistance exercise (post 4hr). The last study was for VL after acute resistance exercise (post 24hr), in which subjects underwent resistance exercise in day 1 and day 3, followed by sample collection in day 2 (here named as 24hr day2) and day 4 (named as 24hr day4).

For normalization, summarization, and removal of batch effect, frozen Robast Multiarray Analysis (fRMA) (*70*) with quantile normalization and robust weighted average summarization was applied to cel raw files, followed by ComBat (*71*) when multiple batches were used (i.e., PBMC datasets). Probes that failed to correspond to unique official gene symbol were filtered out. Log-transformed expression values were further filtered by mean expression and the median expression values were selected from the same gene IDs derived from different probes.

Weighted Gene Co-expression Network Analysis (WGCNA)

To construct gene co-expression network, WGCNA (*47*) was used with in-house R scripts. An adjacency matrix was created by calculating the pairwise gene expression correlation across samples. It is important to confirm that the calculated network displays approximate scale-free topology (*72*). To this end, an adjacency function parameter that leads to sufficient scale free topology fit ($R^2 > 0.8$) was selected based on the result of the pickSoftThreshold function (*57*). We focused on scale free topology fit values rather than exponentially truncated power law in which 2 parameters are involved and often results in overfitting irrespective of the adjacency function (*57*).

To identify the modules in which highly co-expressed genes were grouped, average linkage hierarchical clustering based on topological overlap matrix and dynamic hybrid tree cutting algorithm were applied (*122*). The highly correlated modules (correlation of eigengene: $r > 0.90$ for PBMC, $r > 0.85$ for acute resistance post 4hr) were merged. Notably, despite that there were two time points in acute endurance exercise (pre/post 2.5hr and 5hr), I considered them as one time point (i.e., combined GS values and kME was not affected) because gene expression patterns were almost undistinguishable between post 2.5hr and 5hr.

Network Visualization

For gene co-expression network visualization, I used the modules that showed significant correlation ($p < 0.05$) to exercise traits (either Pre-0min, Pre-1hr, or 0min-1hr) both at eigengene and module levels and also showed significant pathway enrichment in InnateDB (*64*) with hypergeometric algorithm and Benjamini Hochberg *P*-value correction in which pathways in KEGG (*123*) were only focused. Top 1000 strongest edges for each selected module (blue, green, magenta, red, turquoise) were then visualized by Cytoscape (*124*).

Module Preservation Analysis

Different preservation status was determined based on $Z_{summary}$ (strong preservation: $Z_{summary} > 10$, moderate preservation: $6 < Z_{summary} < 10$, weak preservation: $2 < Z_{summary} < 6$, no preservation: $Z_{summary} < 2$). Z score-based analysis measures to what extent the observed network differs compared with the randomly sampled background network. $Z_{summary}$ is the combined statistics of $Z_{density}$ and $Z_{connectivity}$. $Z_{density}$ measures whether the

network module is denser than randomly sampled background, whereas $Z_{connectivity}$ examines whether hub gene status in the module is preserved between the reference and the test network. However, $Z_{summary}$ is often affected by module size. To avoid such bias, I also took into account of medinRank preservation, which measures relative preservation in the module and independent of the module size.

Hub Gene Preservation

To investigate hub gene dynamics, three different hub gene groups were selected. There were: 1) hub genes in the most preserved module (turquoise in PBMC), 2) hub genes in the least preserved module (magenta in PBMC), and 3) least hub genes in any module of PBMC. To select the hub genes in the turquoise module, genes with kME > 0.8 and GS > 0.5 were first selected. The genes were ranked by the order of the absolute values of GS multiplied by kME. Top 20 genes were then selected as hub genes. To select hub genes in the magenta module, genes with kME > 0.7 and GS > 0.5 were first selected. Because the magenta module is much smaller than the turquoise module in size, all the genes (13 genes) that met the criteria were selected as hub genes. The least

hub genes indicate the genes with GS < 0.1 and kME < 0.3 to any module in PBMC (n = 284, that still remained connected to a gene in other conditions).

To examine the status of these selected genes in skeletal muscle, the GS and kME values of these genes in a skeletal muscle condition were bi-plotted. Differently from the connectivity preservation analysis (i.e., $Z_{connectivity}$) which measured whether hub gene status (connectivity to the same module members) in a specific module is preserved in another condition, here I focused on the highest kME, each selected gene has irrespective of the module members in the reference network (e.g., turquoise module members in PBMC). This allows us to investigate whether the genes playing a role as a hub in one condition are naturally hub whatever the condition is or are nurtured hub depending on conditions. As for statistics, ANOVA with TukeyHSD was used.

Preservation analysis within Green Module Network in Acute Resistance Exercise (Post 4hr)

The genes registered in three KEGG pathways (TGF signaling, Insulin signaling, and

MAPK signaling) were obtained from KEGG database (*65*). Two distinct networks, one preserved between acute resistance exercise (post 4hr) and acute resistance exercise (post 24hr) and the other preserved between acute resistance exercise (post 4hr) and acute endurance exercise (post 2.5hr and 5hr), were detected based on: 1) the two node genes of an edge were both present in the three KEGG pathway genes and 2) the edge was commonly observed in both of the conditions. The networks were visualized by Cytoscape (*124*).

The sub-network (*TNFRSF1A, MYC, IL1R1, GADD45A, GADD45B, GADD45G, CRK,* and *NR4A1*) preserved between acute resistance exercise (post 4hr) and acute resistance exercise (post 24hr) was focused and the connected genes to the sub-network in acute resistance exercise (post 24hr) was examined. Top 500 strongest edges were visualized by Cytoscape (*124*). Focusing on absolute GS values of the genes associated with the top 500 strongest edges, Kolmogorov-Smirnov test was used to compare the difference of GS values between acute resistance exercise (post 24hr, day2) and (post 24hr, day4).

III. RESULTS AND DISCUSSION

III-i Modification and Application of Ribosome Profiling to Investigate the Translational Regulation in Mouse Working Muscle

RESULTS

Modification of Ribosome Profiling Using Cultured Cells

I modified the original ribosome profiling protocols (*8, 67*) to best suit for the sequencing library preparation protocols for Ion PGM next gnegation sequencer. In the original protocols, extracted RPF undergo 3´ dephosphorylation, linker/adaptor ligation, and circularization for PCR amplification (Figure 1). However, the sequence library

preparation for Ion PGM sequencer requires adaptor ligations at both ends (3′ and 5′) of the insert without circularization. Therefore, I conducted 3′ dephosphorylation and 5′ phosphorylation simultaneously (Figure 1, details are described in the Materials and Methods).

To confirm that the modified protocols can capture ribosome footprints, RAW264 macrophages were first used as an experimental model. As a control of ribosome profiling, standard mRNA-Seq was also carried out. The results showed highly reproducible sequenced read counts both in mRNA-Seq and ribosome profiling (Figure 2). Pearson correlations of two independent biological replicates in mRNA-Seq were R^2 = 0.97 and R^2 = 0.96, in basal and LPS-stimulated conditions, respectively (Figure 2A,B). In ribosome profiling, two independent biological replicates exhibited high reproducibility, R^2 = 0.97 and R^2 = 0.99, in basal and LPS-stimulated conditions, respectively (Figure 2C,D). Importantly, ribosome profiling shows unique features, called triplet periodicity that is not observed in mRNA-Seq (*8, 67*). In the current study, striking triplet periodicity of sequenced reads was found only in ribosome profiling, but

not in mRNA-Seq (Figure 3), clearly indicating that the modified ribosome profiling successfully obtained the dynamics of translating ribosomes.

Application of the Modified Ribosome Profiling to Mouse Skeletal Muscle

I next applied the modified ribosome profiling and a standard mRNA-Seq to mouse working muscle to investigate the effect of acute endurance exercise on the translational regulation in gastrocnemius. Similarly to RAW264, high reproducibility was observed in skeletal muscle under both of the conditions, basal resting condition (Basal) and immediately after acute endurance running (Exercise). The Pearson correlations of mRNA-Seq were R^2 = 0.98 and R^2 = 0.97 in Basal and Exercise, respectively (Figure 4A,B). Those in ribosome profiling were R^2 = 0.95 and R^2 = 0.84 in Basal and Exercise, respectively (Figure 4C,D). Critically, footprints of ribosomes exhibited unambiguous triplet periodicity only in ribosome profiling, but not in mRNA-Seq (Figure 5), indicative of successful capture of the ribosomal profiles of mouse skeletal muscle in Basal and Exercise.

I next determined a minimum expression threshold for the reliable and robust quantification of transcription and translation. The threshold can be determined by measuring the reproducibility of the independent biological replicates in ribosome profiling and comparing the variation with that predicted by counting statistics (*8, 73*) (Figure 6). As for the genes with less aligned reads, leading to lower statistical power, the variation predicted from binominal partitioning dominated the variation seen in the biological inter-replicates (i.e., the variations predicted by counting statistics and in the inter-replicates were similar) (Figure 6). On the other hand, when sufficient amount of sequenced reads was available (> 125 RPM), the effect of binomial partitioning weakened and the variation in the inter-replicates was larger than that from counting statistics (i.e., other sources of variation, such as biological differences, became larger) (Figure 6). Therefore, to avoid the major effect of counting errors, an expression threshold, 125 RPM, was selected for downstream analyses.

Targeting the genes more than 125 RPM (n = 1,011), principal component analysis

was conducted to examine the global characteristics of transcriptional and translational profiles (Figure 7). There were striking differences observed between transcription and translation as well as Basal and Exercise. Interestingly, clear distinction between transcriptional and translational profiles was found even in Basal as well as Exercise, suggesting that majority of translation and transcription can be regulated separately under both conditions

To narrow down the focus from a global view to individual genes, I identified differentially expressed genes at transcriptional and translational levels. To this end, I normalized transcriptional and translational expression values using a statistical model, in which a negative binomial distribution of sequenced read counts is estimated, and further adjusted the expression bias by applying Trimmed Mean of M (TMM) adjustment (*63*) (Figure 8). Acute endurance exercise induced differential expression ($P < 0.01$) in 26 and 37 genes at transcriptional and translational levels, respectively. Focusing on these genes, pathway enrichment analysis showed the statistically significant enrichments in the pathway associated with amino acid metabolism both in

transcription and translation (Table 1,2). In translation but not in transcription, enrichments were found in the pathways related to innate immunity, generic transcription, ion channel transport, phagosome, and glucose metabolism (Table 1, 2).

From the differentially translated genes, I further focused on Slc25a25, a mitochondrial protein residing on the mitochondrial inner membrane and having a critical role in maintaining ATP homeostasis (*74, 75*). This is due to the facts: 1) Slc25a25 is one of the most differentially regulated genes at translational level, 2) Transcriptional and translational profiles of Slc25a25 showed a significant difference, 3) Mice without Slc25a25 gene decreases endurance performance during treadmill running (*76*), and 4) The effect of exercise, both endurance and resistance exercise, on Slc25a25 expression has never been studied. Although mRNA-Seq showed a moderate increase in Slc25a25 mRNA by approximately 4 fold, ribosome profiling exhibited a much greater raise in translation by approximately 18 fold (Figure 9). These results suggest that acute endurance exercise in mouse can regulate Slc25a25 abundance predominantly by translational regulation rather than that of transcription.

The mTOR signaling cascade is widely recognized as a translational regulator. Activation of mTOR cascade can promote translational initiation and therefore increase translation levels. Given this, one possible speculation is that Slc25a25 translational up-regulation is mediated by mTOR even though previous studies have reported the suppressive effect of endurance exercise on mTOR signaling. To see whether the effect of endurance exercise on mTOR was negative or positive for translation, I focused on TOP motif mRNA. TOP motif is represented by pyrimidine rich sequences immediately after the 5′ cap of mRNA (*77*) and the TOP motif mRNAs were functionally enriched in ribosome/translation (Table 3). It is known that TOP motif mRNAs are highly prone to mTOR inhibition and easily loose the translational efficiency (*77*). Thus, analyzing the translational changes in TOP motif mRNAs make it possible to infer whether mTOR signaling is suppressed and has a negative impact on translational efficiency. In the current study, the translation levels of known TOP motif genes decreased comparing with that in the other genes (Figure 10). This result agrees with the previous studies reporting the endurance exercise-induced decrease in overall protein synthesis (*21, 78*).

Considering the result and that Slc25a25 mRNA does not contain TOP motif, the promoted Slc25a25 translation could be mediated by other factors rather than mTOR signaling cascade.

Because the extent in the increase of Slc25a25 translation exceeded that in the transcription, the extent in the raise of Slc25a25 protein abundance would be much greater than that predicted by its transcriptional regulation. To confirm that Slc25a25 protein abundance is predominantly regulated by translation level rather than transcription, I used western blotting and quantitative real time PCR to measure Slc25a25 protein and transcript levels, respectively (Figure 11). Intriguingly, minor increase (approximately 2 fold increase) was observed in Slc25a25 protein abundance immediately after and post 1 hr of exercise compared with those in translation (approximately 18 fold increase immediately after exercise) and transcription (roughly 7 fold and 4 fold increases in immediately after exercise and post 1 hr of exercise, respectively). These results imply the influence of post-translational regulation, such as protein degradation.

To explain the difference between the translation and protein levels, key players of post-translational regulation, especially protein degradation, were examined. Atrogin1 and MuRF1, well-recognized regulators in muscle protein degradation, were examined. Mmp2 and Nrf2 have been known as critical players in cation-mediated proteolysis and proteasome-mediated protein degradation, respectively (*79, 80*). Given that Slc25a25 protein is a mitochondrial protein, genes regulating proteolysis in mitochondrial, Yme1l1, Immp2l, and Pmpca, were also investigated. Yme1l1 is located in mitochondrial inner membrane and functions as ATP-dependent protease. Immp2l and Pmpca are both responsible for presequence cleavage, though Immp2l is present in inner membrane and Pmpca is found in mitochondrial matrix (*81*). I observed that a single bout of exercise immediately increased expression of Nrf2 (approximately 3 fold increase) at statistically significant level and Mmp2 (approximately 2 fold increase) not statistically significant, though Atrogin1 and MuRF1 showed decreased tendencies but not at statistically significant levels (Figure 12). Although the mitochondrial genes, Immp2l, significantly decreased its expression at 1 hr after exercise, the expressions of

Pmpca and Yme1l1 increased at 2 hr post exercise (Figure 12).

To further examine the association strength among Mmp2, Nrf2, Slc25a25 mRNA, Atrogin1, MuRF1, Immp2l, Pmpca, Yme1l1 and Slc25a25 protein, I conducted a correlation analysis and multiple regression analysis (Figure 13). Interestingly, contribution of Slc25a25 mRNA to Slc25a25 protein abundance was weak throughout the 3 different time points (Figure 13). On the other hand, multiple regression analysis suggested that Nrf2 and Immp2l had negative effects on Slc25a25 protein abundance, though Pmpca was positively associated with Slc25a25 protein abundance (Figure 13)

DISCUSSION

I conducted a genome-wide translational analysis using ribosome profiling to investigate the effect of acute endurance exercise on mouse gastrocnemius. However, there are some limitations in the current studies. One is that we could measure the

global translational profiles only immediately after the exercise. Moreover, due to relatively smaller total read counts I obtained, I could analyze only highly translated or/and expressed genes. Therefore, I could miss many other intriguing regulations that are outside of the current time-courses or below the detection threshold in this study. Another limitation is that ribosome profiling cannot discriminate the translational profile of many mRNAs with monosome (single ribosome on mRNA) from that of single mRNA with polysome (multiple ribosomes on mRNA). Translation of mRNA can be upregulated by either increasing the numbers of translating ribosomes on mRNA or increasing the numbers of mRNAs loaded with a ribosome (or combination of both). However, ribosome profiling cannot distinguish such differences. Therefore, I cannot discriminate whether the currently observed translational upregulation/downregulation were derived from increasing/decreasing ribosomes on mRNA or mRNAs with ribosomes.

Overall, the current study revealed a remarkable distinction between transcriptional and translational profiles even under a basal resting condition. At translational levels, as

other studies have reported that acute endurance exercise could decrease mTOR signaling and therefore reduce protein synthesis (*21, 22*), I consistently observed that TOP-motif genes, prone to the inhibition of mTOR signaling, reduced the translational efficiency. Acute endurance exercise induced dynamic changes both in transcription and translation but with larger extent in translational profiles, exemplified by Slc25a25. Further analyses suggest that transcriptional regulation of Slc25a25 little contributes to Slc25a25 protein abundance but that both translational upregulation and protein degradation could be the key to maintain Slc25a25 protein abundance in the conditions without and early stages after acute endurance exercise. Although the mechanism for the upregulated Slc25a25 translation is unclear, given the decreased translational efficiencies in the TOP-motif genes, mTOR signaling cascade seems not to be involved in upregulating Slc25a25 translation immediately after acute endurance exercise.

Highly enhanced Slc25a25 translation might be traded off by Nrf2-mediated enhanced proteolysis. The current data showed that acute endurance exercise enhanced Slc25a25 translation by approximately 18 fold. However, the corresponding protein

abundance increased only by approximately 2 to 2.5 fold. This huge discrepancy might be explained by promoted proteolysis activity mediated by Nrf2. Nrf2 promotes the gene expression of the components of proteasome complex, including 20S, 19S, and 11S (*80*). Increasing cytosolic Nrf2 level is reported to be necessary to increase proteasome activity (*82*). In the current study, Nrf2 expression was significantly increased by acute endurance running. As the regression analysis suggested, upregulated Nrf2 might have a role in cancelling the elevated translational levels of Slc25a25.

The data suggested that Pmpca and Immp2l could have positive and negative impact on Slc25a25 protein abundance, respectively. Both Pmpca and Immp2l are mitochondrial proteins responsible for presequence cleavage, though their functions are different. When nuclear-encoded mitochondrial proteins are imported into mitochondria, Pmpca cleaves them to be properly folded and functionally matured (*83, 84*). Given that nuclear-encoded Slc25a25 protein has to be processed to properly localize in inner membrane without aggregating or improperly degraded, Pmpca could be responsible for Slc25a25 protein maturation without degradation, supporting the positive effect of

Pmpca on Slc25a25 protein abundance. On the other hand, Immp2l, located in inner membrane, has specific target substrates and the defect in Immp2l can lead to increased mitochondrial ATP levels (*85*). Considering that the loss of Slc25a25 results in decreased mitochondrial ATP levels (*76*) (i.e., the phenotype is negatively correlated with that of Immp2l deficiency), as suggested in the current analysis, Immpl2 might have a negative influence on Slc25a25 protein abundance.

To the best of my knowledge, this is the first study to investigate the effect of exercise on genome-wide translation of skeletal muscle and to focus on the translational regulation of individual genes. Translational profiles significantly differed from those of transcription even under the basal state. The results suggest that acute endurance exercise can enhance the translation of Slc25a25. However, multiple linear regression analysis also suggests that Slc25a25 protein degradation may also have a role in maintaining Slc25a25 protein abundance especially early after acute endurance exercise. Because of decreasing running cost of next generation sequencer, it is now easier to apply genome-wide analysis, which enables us to examine translational regulation of

individual genes. Some of them can be largely regulated by the intensively characterized mTOR signaling cascades. However, the other genes can be mediated by many other known and unknown translational regulators, such as mRNA secondary structure and associated proteins, microRNA, translation speed, or translational stall. I believe that more focus on individual translational regulation would shed novel light on underlying mechanisms of muscular adaptation to exercise.

III-ii. Genome-wide Analysis of Translational Regulation: Alternative Translation

RESULTS

Validation of the currently developed analytical pipeline

Translational regulation is something much more than the simple changes in translational efficiency. Translational dynamics is able to generate multiple protein

species from a single mRNA, including N-terminal extension/truncation, and frameshift. To my knowledge, this is the first time to conduct a genome-wide screening to seek for such polycistronic features in RAW264 and mouse skeletal muscle. The aim of the current analysis is to investigate whether such polycistrons can be found and if any, external stimulus, such as acute inflammation or exercise, can alter the polycistronic expression. To this end, I first conceived a screening methods, in which focus was on the distribution of sequenced read densities within the ORFs in each coding frame, including annotated mORFs and other ORFs that are considered not to be protein-coded. I then visually screened the candidate genes and the read distribution to discover N-terminal extension/truncation and/or frameshifts. To validate the current analytical pipeline, RAW264 data was first analyzed and previously identified alternative translations in mouse were confirmed, followed by screening of mouse working muscle.

In RAW264, previously recognized alternative translations in mouse were successfully captured (Figure 14). For example, there are three uORFs in activating transcription factor 4 (Atf4) (*28*), and N-terminal extension in *Swi5* recombination

repair homolog (yeast) (Swi5) (*28*). As for Atf4 in RAW264, two of the three uORFs were initiated from upstream AUG start codons. One of the two AUG uORF generates a very short polypeptide with 3 residues in length. One of the three uORFs, however, was translated from non-AUG start site, consistent to the previous observation (*28*). In a case of Swi5, uORF was merged with the mORF, producing a long N-terminal extension as the previous study suggested (*28*).

Novel alternative translation in RAW264 macrophages

Previously unknown alternative translations were also discovered. I found a frameshifted uORF overlapping with mORF, encoding alcohol dehydrogenase (Figure 15). Most translating uORFs are short in length (*86, 87*), whereas the newly identified uORF encodes unexpectedly long protein (101 residues). Triplet periodicities for this new uORF and mORF were clearly distinguished upstream and downstream of the overlapping region (Figure 15), supporting the synthesis of the newly identified long uORF. Interestingly, blasting the uORF showed 40 % amino acid sequence identity with the same gene in other species, a part of alcohol dehydrogenase sequences in

Gluconacetobacter diazotrophicus (Table 4). This is surprising because frameshifted ORFs normally produce completely different protein compared with the corresponding mORF. However, the newly identified frameshifted uORF showed a striking similarity with its mORF in other species. Also, other newly identified long translating uORFs (> 50 residues) were all matched with their downstream mORF orthologous at significant identity scores (50 ~ 82 %) (Table 4), implying that long translating uORFs may be byproducts of the evolutional process of mORF as well as the possibility of their own functionality.

As for N-terminal extension/truncation, a truncated isoform and a simultaneous frameshifted polypeptide were detected (Figure 16). For example, there was a truncation in Ras-Related C3 Botulinum Toxin Substrate 2 (Rac2) (full length = 21.4 kDa), playing multiple roles in development and maintenance of stem cells, B cells, or T cells (*88-91*). Rac2 N-terminal truncated translational isoform (deletion of 45 residues) (16.8 kDa) was confirmed by western blotting (Figure 16), in which I used an antibody with an epitope located in the C terminal (143rd to 192nd residues). In addition, the downstream

translational start site was perfectly conserved throughout the different species (Figure 17). Importantly, for a downstream translational start site to be translated, weak Kozak sequence of the upstream annotated start site is critical (*92*). The weak Kozak sequence is also present in different species (Figure 17), suggesting the likeliness that the truncated Rac2 may also be conserved in other species. The truncated region contains some parts of the core functional domain of Rac2, such as GTP/Mg^{2+} binding site, G1 box, G2 box, and Switch I region, implying that synthesizing the truncated Rac2 isoform could have a distinct function and may antagonize the functional Rac2. Taken together, the current approach successfully discovered the previously known and unknown alternative translations, including uORFs and frameshifts in RAW264.

Novel alternative translation in mouse skeletal muscle

Applying the screening method to mouse skeletal muscle also discovered novel polycistronic transcripts and their translational byproducts (Table 5). Even though only highly expressed genes were focused, 2 truncations and 8 uORFs were successfully detected. The N-terminal truncated genes were lactate dehydrogenase A (Ldha) and

triosephosphate isomerase 1 (Tpi1), each of which lacked in its N-terminal by 8 % (29 residues) and 17 % (50 residues) of the full-length protein (Figures 18, 19). However, the truncated polypeptides were not in the part of the conserved functional domain. As for the newly identified uORFs, all were short in length. There was no difference in these alternative translations between with and without exercise stimulus. In other words, taken together with that these transcripts are highly translated mRNAs, these truncations and uORFs can also be constantly generated at high translation levels in mouse skeletal muscle.

DISCUSSION

Dynamic alternative translation was successfully identified in the current screening both in RAW264 and mouse skeletal muscle partly due to their underestimated importance. Because these alternative products are not annotated and previously uncharacterized, its functional potential is of interest. One of the reasons that ones have

missed such alternative translation derived from multiple ORFs in a gene can be that initially annotated protein coding regions of genes were manually assigned based on several criteria, such as minimum length of ORFs. This resulted in a bias that small ORFs, including uORFs, are not functional or only some biological background noise.

Recently, however, the significance of such small polypeptide has drawn attention. In 2013, it was revealed that small polypeptides (< 30 residues in length) have critical effects on cardiac muscle contraction and the lack of the small proteins leads to cardiac dysfunction (*93*). Considering the biological importance of such small polypeptides by which cardiac contraction is constantly maintained, they need to be constantly expressed to maintain its indispensable importance no matter what the surrounding conditions are. According to the currently observed results in mouse working muscle, the constantly translated uORFs at high expression levels are of interest because they were abundantly and constantly present in the different conditions, including the basal and after acute endurance exercise. However, functional screening of these uORFs can be difficult due to the tendency that small polypeptides conserve their functionality based on the

confirmation rather than peptide sequences (*93*), making it difficult to simply comparing and blasting peptide sequences to infer the molecular function.

N-terminal truncations were identified both in RAW264 macrophages and mouse skeletal muscle. In RAW264, the deleted region contains the conserved domain of the protein, Rac2. Taken together with the weak Kozak sequence preserved across species, it can be suggested that the truncation can be evolutionally conserved. In contrast, there was no functionally important domain in the truncated regions of the newly identified alternative translations in mouse skeletal muscle. It cannot exclude the possibility that the deletion of N-terminal may alter the protein conformation, though its physiological influence might be negligible compared with that in Rac2. Taken all, in the current analysis, I successfully identified novel alternative translations in RAW264 macrophages and mouse skeletal muscle using originally generated sequencing data. Although the functional roles of these byproducts remain unknown, it is intriguing to investigate whether they have any biologically meaningful functions or are just only cellular background noise.

III-iii. Genome-wide Analysis of Translational Regulation: Dynamic Change in Translation Speed of Ribosomes

RESULTS

Reduced bias in ribosome density by binary counts

To conduct genome-wide analysis of translation speed, I used ribosome density data in RAW264 macrophages and mouse skeletal muscle. For a robust measurement of ribosome density to estimate translation speed, I established a novel analytical method, named binary count approach, in which aligned read counts at a base position of a gene are converted to 0 or 1 binary count. To validate binary count, unique characteristics of ribosome profiling data were examined. I compared the RPF count patterns between the raw read counts and binary counts by metagene analysis (Figure 20A). Both raw and

binary counts showed the strongest peak at the position 12 nt upstream of the start site as previous studies reported (*8, 67*). Because ribosomes move 3 nt by 3 nt, it is also known that RPF show 3 nt periodicity (*8, 67*). Similarly in the both raw and binary count plots, 3 nt periodicity of ribosome density was observed in which the highest peak indicates the main coding frame at each codon (Figure 20A). However, the stability in the main coding frame showed striking differences between the raw and binary counts (Figure 20A). The binary count was more stable than the raw count was. To statistically confirm this, I calculated the cumulated CV of the ribosome counts in the main coding frame (Figure 20B, raw and binary densities colored in black and red, respectively). Kolmogorov-Smirnov test showed significantly reduced CV in the binary counts, suggesting that binary count can stabilize the accumulated ribosome counts in the main coding frame without eliminating the global features uniquely present in ribosome profiling.

Translational arrest core (TAC) sites detected by binary counts

Taking advantage of the binary counts, I investigated the effect of each of 20 residues

on ribosome density in RAW264 macrophages. To this end, ribosome density on a specific gene sequences in a 105 nt window was observed. Importantly, in a 105 nt sequence window, there is only one codon (residue) of interest, which makes it possible to detect the effect of a single residue of interest on ribosome density, and therefore presumably translation speed (Figure 21). If a residue of interest causes significant accumulation of ribosomes (i.e., slow translation speed), a ribosome density peak can be observed depending on where the residue stalls a ribosome (Figure 21). I assumed that different residues would have distinct effects on ribosome density patterns. Therefore, ribosome density peaks can be varied. To first observe a global picture of 20 different ribosome density peak patterns, I carried out a metagene analysis. To avoid sequence bias and to normalize ribosome density, cumulated RPF counts (raw or binary) at each relative base position in the defined sequence window were divided by cumulated mRNA-Seq counts (raw or binary) at each corresponding relative base position. I then superimposed the 20 normalized ribosome density patterns onto one plot, each of which represents the effect of single codon of interest on ribosome density and assumingly on translation speed. When the raw density was used, large density variations in the main

coding frame were observed (Figure 22). On the other hand, when binary count approach was applied, the variations were minimized and striking translational arrest core (TAC) sites, position 0 to +20 nt relative to ribosomal A site, were identified (Figure 22). The TAC sites composed of A/P sites and extra 5 residues. These results clearly indicate the robustness of binary density and that the binary approach allows us to discover highly conserved TAC sites. Although the density peaks in TAC sites could be observed in the raw density plots, the background noise was huge to the same or even larger extent, which masked the peaks in TAC. Of course, I cannot completely exclude a possibility that the large variations in the raw density might be true positive signals. However, given the disturbed 3 nt periodicity in the raw density plots (Figure 22) and the significantly lowered CV in the binary density (Figure 20B), it is more plausible to consider that the binary density better captured the translational dynamics than the raw density did.

Differentially regulated translational stalls in a stimuli-dependent manner

Using the newly developed binary count method, I compared the effect of 20

different residues (Figure 23-29) and corresponding 61 codons on ribosome densities during basal and LPS-induced acute inflammation state. This analysis enables one to understand to what extent the effect of residue or codon predominates in translational stalls at each position (e.g., within ribosomal exit tunnel, A- and P-sites, or upstream of the coding center) as well as whether acute inflammatory stimuli can render translational stalls in RAW264 macrophages. Interestingly, there were highly correlated ribosome density patters at A- and P-sites as well as the positions 1 and 2 codons away from the coding center irrespective of basal or inflammatory condition (Table 6). However, inflammatory stimuli lowered correlations at the positions 3 and 4 codons away from the coding center (Table 6). To further reveal which codons caused translational pausing in the different conditions, differential translational pausing induced by acute inflammation was examined. To this end, I first confirmed that ribosome density measurement was highly reproducible between two independent biological replicates in both conditions (Figure 30, 31). Ribosomal densities induced by each codon in the basal and LPS stimulation were then bi-plotted (Figure 32A). Surprisingly, different codon-position pairs of 12 and 20 increased or decreased

ribosome densities, respectively under the LPS-induced acute inflammatory condition in RAW264 macrophages. However, acute endurance exercise had little effect on ribosome density (Figure 32B).

The effect of LPS-induced alteration in translational stalls on nascent protein folding

Altered translation speed can render nascent protein conformation. It has been reported that artificially modified genetic codon usage resulted in differential translation speed and protein structure (*33, 94*). However, there has been no report addressing that actual physiological change can alter translation speed and protein structure. Thus, I further investigated whether the currently observed global changes in translation speed could also be found in individual gene level and the dynamics in translational pausing could render nascent protein conformation. From the codons that showed significant changes in translational stall during acute inflammation, I focused on a codon, GAC. This is because acute inflammation affected RPF length (Figure 33). As indicated in a previous study (*95*), ribosomal stall can change the corresponding RPF length.

Therefore, GAC was a primary target codon. I then selected a gene, Calreticulin, exhibiting a striking alteration in ribosomal stall at GAC codon (Figure 34). To further validate that the observed differential ribosomal stall in Calreticulin can affect its protein structure, I conducted a trypsin assay. Trypsin assay is based on the idea that differential protein structure should expose varied polypeptide sequences to trypsin and therefore lead to distinct susceptibility to trypsin digestion. Trypsin assay, targeting Calreticulin, clearly exhibited distinct digestion patterns (Figure 35). In contrast, GAPDH, showing much smaller dynamics in ribosome stalls, failed to show the difference (Figure 35). These results suggest for the first time that acute inflammatory stimuli can render ribosomal stalls and further affect corresponding protein conformation.

DISCUSSION

Novel analytical method, binary count, was applied and successfully reduced

background noise of ribosome densities. Because ribosome profiling is robust and sensitive enough to capture a unique translational feature in a single gene, such as frameshift (*8, 28, 46*), diverse characteristics of different genes, including distinct base composition and gene expression levels, could prevent a high resolution analysis of translational stalling. For example, highly expressed genes predominate the total sequenced reads, which subsequently leads to that total ribosome densities reflect only the small fraction of such genes. In contrast, low expressed genes little contribute to the observed global density features. The biased ribosome density might hide the potentially interesting global characteristics of translation and make it difficult to identify where in ribosome translational arrest occurs. Binary count approach made it possible to examine ribosome density at a high resolution by avoiding such bias.

One of the strength of the current study is that I made it possible to measure translational stall at a high resolution and identified exactly where in ribosome translational stall occurred. The accuracy and reliability of the positional assignment by binary count is strongly supported by previous studies reporting that translational stalls

at P and A sites can be caused by Pro and Glu, respectively (*28, 41*). Consistently, I also identified Pro- and Glu-induced strong translational stalls at P and A site, respectively. Moreover, as for stalling within exit tunnel, cryo-EM structural basis *in vitro* analysis has recently shown that the insertion of Met residue at a position of a polypeptide chain can cause by far the most significant translational stall compared to when the other residues were substituted (*37*). This observation perfectly agrees with my current finding in which Met residue significantly arrested translation at the position corresponding to the 4^{th} residue away from the P-site. Given these, it is clearly evident that the binary count approach makes it possible to detect transient translational stalls and identify exactly where the stalls occur at a high resolution. In addition, these explicit agreements between the previous studies and the current results strongly reinforce the validity of the binary approach and the reliability of the assignment of density peak positions.

Intriguingly, significant contradiction was observed between the ribosome densities induced by single residue and single codon. In theory, as far as the codons encode the same residue, the effect of codon on ribosome density or translational stall should be the

same if it is the nascent residues that influence translation. The current observation was partially not the case. Different codons, encoding a same residue, often resulted in different ribosomal densities, suggesting that not all codons affected translation via its residue. Alternatively, they could induce translational stall as a cis-element or in a codon sequence-dependent manner. As a cis-element effect, mRNA secondary structure is a potential candidate that cause the currently observed ribosomal density peaks. Given the nature of secondary structure of mRNA (e.g., hairpin structure), such cis-factor should hamper ribosomal movement downstream of the RPF 3′ end. However, most of the ribosomal peaks were observed from near the A site to 5 residues (15 nt) away from the P site (RPF 5′ end side). This indicates that the codon resides within the ribosomal complex when the codon triggers accumulation of ribosome or translational stall, implying the unlikeliness of secondary structure-induced stall of ribosome at least in the current observation. Alternatively, as suggested in anti-Shine-Dalgarno sequence, by which ribosome can stall during translation (*95, 96*), the affinity between codon sequences and ribosomal complex might cause the currently observed ribosomal density peaks and related translational stalls.

LPS-induced alteration of ribosome density and translational speed can imply a novel regulatory mechanism in biology. In this study, LPS but not acute endurance exercise resulted in remarkable changes in ribosomal density peaks, indicating dynamic shift in translation speed. Recently, it has been reported that artificial modification of codon usage could alter translation speed and corresponding protein confirmation and its function (*33, 94*), indicating that change in translation speed can render protein structure and its function. However, no study has reported that normal physiological change can utilize such regulation and actually alter translation speed. To the best of my knowledge, this is the first study to exhibit that change in physiological condition (e.g., acute inflammation) can alter translation speed, inferred by the changes in ribosome density peaks, and importantly, can induce distinct nascent protein conformation via locally altered translation speed.

III-iv. Genome-wide Analysis: Gene Co-expression Network Analysis

RESULTS

Another genome-wide approach is co-expression network analysis, which can complement the conceptual lack (e.g., gene network and hub genes) in the conventional genome-wide analysis as I described above. Gene co-expression network analysis is particularly useful to understand a complex interaction (e.g., effect of exercise) from a systematic perspective. Because different types of exercise have sophisticated effects on different tissues at different time-courses, it is intriguing if one can understand the underlying regulation and the dependency/independency of exercise types, tissues, and time-courses. The aim of the current analysis is to understand to what kinds of network-level dynamics can be preserved across different exercise types, time-courses, and tissues. This can be achieved using the publicly available microarray datasets of human and analyzing gene co-expression network preservation.

Research Design for the Analysis of Gene Co-expression Network Preservation across Exercise Types/Time-courses/Tissues

I conducted two different sets of preservation comparisons (Figure 36). First comparison defined acute endurance exercise-induced peripheral blood mononuclear cell (PBMC) gene co-expression network as a reference network. The other conditions in skeletal muscle vastus lateralis (VL), namely acute endurance (pre/post 2.5hr and 5hr), acute resistance (pre/post 4hr), and acute resistance (pre/post 24hr), were defined as the test networks (Figure 36 top). In the second comparison, the network in acute resistance (pre/post 4hr) was the reference network (Figure 36 bottom).

These two sets of comparison make it possible to compare maximum numbers of network preservation under the limited exercise conditions. Therefore, it was possible to examine: 1) Tissue/time-independent preservation, in which network modules in PBMC are preserved in the other VL conditions, suggesting that the modules respond to exercise stimuli in a tissue/time-independent manner, 2) Muscle-specific preservation, in which modules in acute resistance (pre/post 4hr) are preserved in the other VL

conditions, but not in the PBMC network, 3) Muscle-specific and resistance exercise-specific preservation, in which preservation exists only in the muscle tissue undergoing resistance exercise, and 4) Muscle-specific and early stage-specific preservation, in which preservation is present only in the muscle tissue during early stages after exercise no matter what the exercise type is.

Construction of PBMC Network after Acute Endurance Exercise

To investigate whether there is tissue-independent network preservation between PBMC and VL, I first constructed PBMC weighted gene co-expression network using WGCNA R package and in-house R scripts. An adjacency matrix of PBMC was first calculated and the highly correlated modules were merged together ($r > 0.90$) (Figure 37). To narrow down biologically interesting modules for preservation analysis, three criteria were used based on the significance of each network module and its functional enrichment: 1) Significant Pearson correlation ($p < 0.05$) between eigengene and an exercise trait (Figure 38), 2) Correlation ($r > 0.40$) between module membership (kME) and absolute gene significance (GS) (Figure 39), and 3) Pathway enrichment analysis

for the module show a statistical significance (Table 7). Eigengene here represents its network module and is the first principal component of the module. The exercise traits in PBMC analysis are composed of “Exercise”, “ExercisePre0min”, “ExercisePre1hr”, “ExerciseTime”, and “Batch” (Figure 38). “ExercisePre0min” and “ExercisePre1hr” indicate the comparison between pre-exercise and immediately after or 1hr after the exercise, respectively. “Exercise” reflects the combined effect of immediate and 1hr after the exercise. “ExerciseTime” compares the two post-exercise time points. “Batch” compares the two independent datasets used in this analysis to confirm that there is no data source bias. kME measures how strong the connection is between a module member (gene) and the eigengene of the module. GS for a gene indicates the correlation between the gene expression patterns and exercise trait. For example, high absolute GS in “ExercisePre1hr” for a gene can be obtained when expression patters of the gene throughout all the samples are increased (or decreased) post 1hr of exercise compared with pre-exercise. Therefore, if a network module shows a high correlation between kME and absolute GS values (i.e., the stronger the connection is, the more the significance of the gene is), the module can be an important and meaningful network

during the systematic regulation in the condition.

Based on the selection criteria (1) to (3), I selected five modules (colored in turquoise, magenta, green, blue, and red) in PBMC. Pathway enrichment analysis revealed unique functional differences among these five modules (Table 7). The magenta and green modules were enriched with immune cell specific functions, such as NK cell mediated cytotoxicity, T cell receptor signaling, and antigen processing and presenting. Due to the closely related functional enrichments, the magenta and green module networks were positioned close together in the visualized network (Figure 40). The relation between functional distance and visualized network distance was also the same for the red and blue modules. Both red and blue modules showed significant enrichment in phagosomal function (Table 7) and these network modules were located close to each other (Figure 40). The turquoise module, however, showed relatively dispersed module network (Figure 40). This could be due to its multi-functional features shown in the pathway enrichments (Table 7). These results suggest that PBMC responds to acute endurance exercise by forming and regulating the two distinct networks: 1) adaptive immunity

derived by the magenta – green modules, tightly coupled to the turquoise module and 2) innate immunity, particularly phagosomal and lysosomal functions, derived by the blue – red modules. Moreover, given the correlations with the exercise traits (Figure 38,39), acute endurance exercise seems to weaken the network functionality of adaptive immunity and enhance that of phagosomal and lysosomal functions in PBMC after 1hr of endurance exercise. Therefore, as a reference network of PBMC in following network preservation analyses, I focus especially on the PBMC (pre/post 1hr of exercise) gene co-expression network, not pre/immediately after exercise.

PBMC Network Module Preservation in VL

I next investigated whether the five PBMC modules were preserved in VL responding to similar or different types of exercise. Network preservation can be measured by the permutation of random sampling of the background network to calculate the noise and $Z_{summary}$ values for each of the reference modules in PBMC. $Z_{summary}$ is the summary statistics of $Z_{density}$ and $Z_{connectivity}$. $Z_{density}$ measures whether the network module is denser than randomly sampled background. $Z_{connectivity}$ measures whether the hub gene

status in the module is preserved between PBMC and VL. $Z_{summary}$, however, often depends on module size. To avoid the module size-induced bias, relative preservation statistics (medianRank) was also calculated, which measures relative preservation only within test networks, meaning that absolute preservation cannot be measured by relative preservation statistics.

Using the Z statistics and relative preservation statistics, I examined PBMC module preservation (Figure 41-43). The turquoise module in PBMC was the most preserved in all the VL conditions (Figure 41, Table 8). Comparing the density and connectivity preservation, the turquoise module was preserved more in connectivity than in density (Figure 42,43). In contrast, the magenta module showed no preservation in any VL condition (Figure 41-43, Table 8). This seems a reasonable outcome given the enriched pathway functions in the turquoise and magenta module. Although the turquoise module has basic and multi-functionalities (e.g., Nucleotide excision repair, Protein export, Proteasome, Pyrimidine metabolism, Basal transcription factor, and RNA transport), the magenta module specializes in adaptive immunity (NK cell mediated cytotoxicity and T

cell receptor signaling). Given this, the distinct preservation status illustrates that a co-expression network module responsible for a highly specific network module (as seen in the magenta) is unlikely to be preserved between PBMC and VL even in a similar exercise condition. However, a module with basic cellular response or functions (as seen in the turquoise) is likely to be preserved between PBMC and VL no matter what the exercise types or time-courses are.

In the case of a specialized innate immunity, lysosome and phagosome, a conditional preservation was observed in the red – blue modules. These modules showed weak but constant preservation (i.e., weakly preserved in all $Z_{summary}$, $Z_{density}$, and $Z_{connectivity}$) in acute resistance exercise (post 24hr). However, in the same resistance exercise condition but different time point (post 4hr), no preservation and marginal preservation were observed in the red and blue modules, respectively (Figure 41). Given such conditional preservation, it is suggested that innate immunity, especially lysosomal or phagosomal function, forms its functional network similar to that in PBMC (post 1hr) under the condition of acute resistance exercise (post 24hr) but not acute resistance

exercise (post 4hr) or acute endurance exercise (post 2.5hr and 5hr).

PBMC Network Hub Gene Status Preservation in VL

Hub genes are the central players in a network module. I was next interested in observing the hub gene status, that is, to what extent the hub genes in one condition can retain their influence to the whole network in other conditions, not only to their preserved module members. Although WGCNA enables one to examine the preservation of hub gene connectivity (Figure 43), it focused on the connectivity between hub genes and their module members in a reference module, not including other genes in the other modules (Figure 44A). Here focusing on the hub status irrespective of the original module members in the reference module (Figure 44B), I examined the dynamics of hub gene status in different conditions. The question I address here is whether hub status is determined in a nature or nurture manner. I speculated that the hub genes were naturally hub and that the hubs of a well-preserved module were likely to keep their hub status in a different condition compared with those of a less-preserved module.

Because hub gene expression patterns show strong correlation with module eigengene and with traits, I defined hub genes as the genes with high kME and high absolute GS. kME and GS describe strength of eigengene-based connectivity and correlation between gene expression patterns and an exercise trait, respectively. Two hub gene groups and one not-hub-gene group were selected: 1) Hub genes in the most preserved module of PBMC (turquoise module), 2) Hub genes in the least preserved module (magenta), and 3) Least hub genes in any module of PBMC network. Surprisingly, all the three different group genes showed significant changes in their hub status across PBMC and VL (Figure 45-47). As for GS, although the least hub genes increased their absolute GS values, the hub genes both in the most preserved (turquoise) and the least preserved (magenta) modules significantly decreased the absolute GS values (Figure 48-50). As for kME values, the least hub genes increased their kME whereas the hub genes both in the turquoise and the magenta modules significantly decreased their kME, except in acute resistance exercise (post 24hr) (Figure 51-53). In acute resistance exercise (post 24hr), both of the hub genes in the turquoise and magenta

did not significantly lose their kME values (Figure 52,53).

Because the hub status of the turquoise/magenta modules remained strong in acute resistance exercise (post 24hr), I further examined whether their associated network functions remained the same in acute resistance exercise (post 24hr) (Figure 54, Table 9). Interestingly, more than 99.9% of all the genes connected by top 1000 strongest connections (top 500 edges connected to the hub genes from the PBMC magenta, top 500 from the PBMC turquoise) corresponded to one single module, a new turquoise module in acute resistance exercise (post 24hr). Moreover, the magenta/turquoise PBMC hub genes shared most of their newly connected genes (62.6 % of magenta, 65.2 % of turquoise) and contributed to significantly enriched cellular functions, namely protein processing in ER and endocytosis (Figure 54, Table 9). This illustrates a striking flexibility of hub gene connectivity because the original turquoise/magenta PBMC modules were distantly related (Figure 40) and they did not share their functions in PBMC (Table 7). These results indicate the flexibility and nurtured characteristics of hub genes. It is further suggested that any gene has a potential to play a key role or lose

its importance in a certain network. Additionally, even two distantly related hub genes with two distinct functions in one condition can keep their hub status in another condition to orchestrate together to construct a shared new network and function.

Muscle-Specific Network Preservation

I next investigated whether there is any muscle-specific network. To this end, adjacency matrix was created and highly correlated modules ($r > 0.85$) were merged for the network in acute resistance exercise (post 4hr) (Figure 55). To find biologically interesting modules, Pearson correlation was calculated between each eigengene and different traits ("AcuteResistance", "MaleFemale", "YoungOld") (Figure 56). As in the previously described PBMC network, to calculate the Pearson correlation, binary values were manually assigned to each trait (i.e., before exercise = 1, post 4hr = 2, male = 1, female = 2, young = 1, old = 2). Inclusion of sex and age traits enabled me to confirm that the significant correlation derived from exercise was not a cofounding of sex or age. For further analysis, I filtered the modules based on the followings: 1) Significant correlation ($p < 0.05$) between eigengene and an exercise trait, 2) Correlation ($r > 0.35$)

between kME and absolute GS (Figure 57), and 3) Pathway enrichment analysis for the module show a statistical significance (Table 10). After filtering, five modules were selected for further analysis. Among the five, the green module showed a strikingly high correlation both at eigengene (r = 0.88) and module member (r = 0.90) levels (Figure 56,57).

Focusing on the five modules, I next examined the preservation status of each module in other conditions including PBMC, acute resistance exercise (post 24hr), and acute endurance exercise (post 2.5hr and 5hr). As described in PBMC network preservation, Z statistics and median rank statistics were used to compare preservation (Figure 58-60, Table 11). Two modules, green and darkgrey, were preserved in muscle tissues, but not in PBMC. Pathway enrichment analysis showed that the green module was enriched with the genes related to transforming growth factor (TGF) signaling, insulin signaling, and mitogen-activated protein kinase (MAPK) signaling and that the darkgrey module was enriched with extracellular matrix (ECM) receptor interaction, focal adhesion, and protein digestion and absorption (Table 10). These results indicate that the green and

darkgrey module networks are preserved in muscle tissues independently of exercise modes or time-courses after exercise.

Exercise Type Specific Preservation

Because the green module was highly correlated with acute resistance exercise (post 4hr) and suggested a critical importance during the adaptation to the exercise stimuli, I depth-analyzed if there was any preservation difference within the green module depending on exercise types or time-courses after exercise. I first focused on the preservation in an exercise mode specific manner. Since the muscle-specific green module was enriched with three KEGG signaling pathways, including TGF, insulin, and MAPK pathways (Table 10), the genes involved in the KEGG pathways were used to find what network in the three signaling cascades was preserved between acute resistance exercise (post 4hr) and (post 24hr), but not in acute endurance exercise. To this end, I extracted the common network connections (edges) between acute resistance exercise (post 4hr) and (post 24hr). I then selected the edges whose node genes were both found in the three KEGG pathway gene lists to focus only on the network within

the pathways. The obtained network was composed of two major sub-networks (smaller and larger) preserved in both conditions (Figure 61-63), which were hardly detected in the network of acute endurance exercise. The network in acute resistance (post 24hr) had two different sets. One is for the VL sampled in day 2, that is, 24hr after the resistance exercise carried out in day 1. The other is for the VL sampled in day 4, that is, 24hr after the repeated resistance exercise in day 3.

Among the acute resistance exercise (post 4hr), (post 24hr, day 2), and (post 24hr, day 4), there were striking differences in GS values in the preserved network, especially observed in the smaller sub-network (Figure 61-63). The smaller sub-network was composed of 8 core genes: *TNFRSF1A, MYC, IL1R1, GADD45A, GADD45B, GADD45G, CRK,* and *NR4A1*. Although the network connections among the 8 genes were well preserved between acute resistance exercise (post 4hr) and (post 24hr), the GS values of the 8 genes changed over the time. In acute resistance exercise (post 4hr) and acute resistance exercise (post 24hr, day 2), most of the GS values were positively correlated with the exercise (7/8 in post 4hr and 6/8 in post 24hr day 2) whereas the GS

values in acute resistance exercise post 24hr (day 4) turned out to be negatively correlated or not correlated with the exercise (Figure 61-63).

There was also dynamic change in the associated network with the 8 genes comparing post 4hr with post 24hr. Interestingly, although the 8 genes were involved in the three cascades (MAPK, TGF, and Insulin) under the acute resistance exercise (post 4hr), these genes formed a new phagosomal network in acute resistance exercise (post 24hr) (Figure 64, Table 12). Moreover, the network genes contributing to the phagosomal function showed high absolute GS values in post 24hr (day 2), though they reduced the absolute GS in post 24hr (day 4) (Figure 65), which is consistent to the GS dynamics of the 8 genes themselves (Figure 62,63). Taken together, these results indicate that green module preservation, especially the tight connection among the 8 genes (*TNFRSF1A, MYC, IL1R1, GADD45A, GADD45B, GADD45G, CRK,* and *NR4A1*), was present only in acute resistance VL but not in acute endurance VL or PBMC. It can be suggested that these 8 genes have an important role in forming phagosomal network under the acute resistance exercise post 24hr (day 2) but not in

post 4hr or with weaken network significance in post 24hr (day 4).

Post Exercise Time-course Dependent Preservation

I next examined the network preservation specifically present during the early stages of post exercise. To this end, I focused on the green module preserved between acute resistance exercise (post 4hr) and acute endurance exercise (post 2.5hr and 5hr). Similar to the previously described exercise type specific preservation, the genes associated with the three pathways (MAPK, TGF, and Insulin) in KEGG were used again to screen for the preserved network between acute resistance exercise (post 4hr) and acute endurance exercise (post 2.5hr and 5hr). The early stage specific network revealed one strongly preserved network between acute resistance exercise (post 4hr) and acute endurance exercise (post 2.5hr and 5hr) (Figure 66,67). In particular, *RHEB – PRKAG2 – PPARGC1A* showed a striking preservation as well as their positively correlated GS values. These networks were observed only in acute resistance exercise (post 4hr) and acute endurance exercise (post 2.5hr and 5hr), but not found in acute resistance exercise (post 24hr) or PBMC.

DISCUSSION

In the current study, I investigated the preservation of the gene co-expression network across different exercise modes, tissues, and post exercise time-courses to understand the network similarities and differences induced by distinct exercise conditions. First focus was on PBMC gene co-expression network construction triggered by acute endurance exercise. The association between the co-expression network and post exercise conditions illustrated that network modules, related adaptive immunity, became attenuated post 1 hr of endurance exercise whereas innate immunity, specifically phagosomal and lysosomal functions, was reinforced. Phagocytosis is a critical innate immune step to clean up pathogens and external cellular debris. Internalized phagosome is fused with lysosome to form phagolysosome, where encapsulated targets are digested (*97*). The currently observed activation of phagosome/lysosome network can be supported by the previous studies, in which phagocytic functions of macrophages

significantly increased after endurance exercise (*98, 99*). As for the decreased activity of adaptive immune network in this analysis agrees with "open window" theory, suggesting exercise-induced temporal immune suppression (*100, 101*). It should be noted that I cannot exclude the possibility that the currently observed network dynamics might be caused by not only gene-expression changes but also by the alteration in monocyte population. Taken together, systematic counter-balancing between innate and adaptive immunity seems to occur in PBMC after endurance exercise.

Genes, responsible for such altered network, can be suggested as hub genes. In the magenta module forming the suppressed adaptive immune network, for example, *STAT4*, *IL2RB*, *PRKCH*, and *GRP56* appears to have an important role as hub genes. *STAT4* is a crucial transcriptional regulator that regulates T_H1 cells (*102, 103*) and IL-2-triggered NK cell activation (*104, 105*). *IL2RB* produces a subunit of IL-2 receptor complex and contributes to both intermediate and high affinity forms of the receptor, critical for IL-2-induced signal transaction (*106*). As indicated by these well-investigated roles in immune system, the hub genes, *STAT4* and *IL2RB*, can exert critical influences on the

immune adaptation to acute endurance exercise. *PRKCH* and *GRP56* were also identified as the hubs in the magenta module. *PRKCH*, a gene of a serine/threonine-protein kinase, has been reported that its polymorphism is associated with immune disorder, such as rheumatoid arthritis (*107*). GRP56 has been recognized as the marker of cytotoxic T and NK lymphocytes (*108*) and that its overexpression leads to NF-kappaB activation (*109*). Despite their unignorable roles in immune system, to my knowledge, their functional importance during exercise adaptation has never been investigated. The current gene co-expression network analysis can suggest that these less characterized hub genes can potentially have a role in modulating the immune network. Considering that the magenta module was involved in adaptive immunity and its network functionality was attenuated after 1 hr of acute endurance exercise in PBMC, it may be that regulation of *PRKCH* and/or *GRP56* may modulate, at least in the part of, "open window" theory after endurance exercise.

Although network analysis suggested that hub genes (e.g., *PRKCH* and *GRP56*) might be a key regulator in the adaptive immunity after 1 hr of endurance exercise, this

may not be true in other conditions. In the current study, preservation analysis of hub genes revealed the flexible nature of hub gene status. Even if a gene that is a strong hub or not hub at all in one condition, the gene can easily lose or gain “hubness” in another condition. Alternatively, as the current result showed, in a case that the original hub gene (in PBMC) remained hub in another condition (acute resistance exercise, post 24 hr), the genes became connected with new genes forming a totally different functional network compared with the original one, further indicating the functional flexibility of hub genes. This feature can be represented by *MYC*. *MYC* is a central regulator in somatic cell reprogramming (*110*) as well as in regulating ribosomal biogenesis (*111, 112*), differentiation (*113, 114*), or apoptosis (*115*), representing a key player with distinct functions in the wide range of cellular conditions. Taken together, the current hub gene analysis indicates flexible hub gene nature. Hub genes can belong to a distinct cellular function when they keep their “hubness” in a different environment or they can possess their “hubness” only in a specific condition.

In terms of the preservation of PBMC network in VL conditions, a specialized

network, as in the magenta, was less preserved. This seems a reasonable outcome because PBMC specializes in immune response and inevitably forms immune-specific gene co-expression network, which is not found in skeletal muscle. On the other hand, a striking preservation was found in turquoise module, enriched in house keeping-like diverse functions, including nucleotide excision repair, protein and RNA processing. The critical roles of the house-keeping or mediators appear indispensable no matter what the cell types or external stimuli are and have a consistent role to mediate an input stimuli and differential outputs depending on the conditions.

Preservation analysis in skeletal muscle co-expression network exhibited a tightly connected network preserved in a resistance exercise-specific manner, which may also have a role in regulating the adaptation to the repeated effect of exercise. The network, composed of *TNFRSF1A, MYC, IL1R1, GADD45A, GADD45B, GADD45G, CRK,* and *NR4A1,* was preserved between acute resistance exercise (post 4 hr) and acute resistance exercise (post 24 hr). Interestingly, despite the long-lasting activation of the network even after 24 hr of 1st round of resistance exercise (post 24hr, day 2), the elevated

activation was totally diminished after 24 hr of 2nd round (post 24hr, day 4). This suggests that the distinct skeletal muscle adaptation to the 1st and 2nd round of resistance exercise might be attributed to the enhanced and decreased expression of these core network genes. This network dynamics was not found by a conventional differentially expressed gene analysis. The original study of the currently used dataset addressed an issue that there were huge individual variations after the 1st round of resistance exercise, resulting in blunting the signs of differentially regulated genes (*116*). In WGCNA, however, the co-expression network was constructed based on the correlations between each pair of gene expression patterns, which was independent of the extent of expression changes. Therefore, the current analysis could capture the similarly co-expressed important gene network avoiding the negative effect of diverse individual differences. Taken together, there was a resistance exercise-specific network present from post 4 hr to at least post 24 hr of exercise, which was formed by *TNFRSF1A, MYC, IL1R1, GADD45A, GADD45B, GADD45G, CRK,* and *NR4A1*. Moreover, the expression patterns of the core network showed a striking distinction between after 24 hr of 1st and 2nd round of resistance exercise, suggesting a link to the differential muscle

adaptation to repeated exercise. Given also that *MYC* modulates many different cellular responses by transcriptionally regulating diverse gene families, differential regulation of *MYC* may be the key for a conversion of resistance exercise stimuli to successful muscle adaptation, such as hypertrophy or resolution of inflammation.

A highly preserved co-expression network between resistance and endurance exercise was also identified at relatively early stages of post-exercise, implying its role for further differentiated muscle adaptation to different exercise modes. This exercise mode-independent preservation was represented by the tight connection among *RHEB* – *PRKAG2* – *PPARGC1A*. Interestingly, these genes are involved in the regulation of mTORC1 (mammalian target of rapamycin complex 1). The mTORC1 is a well-known translational initiation activator. *RHEB* is a member of small GTPase Ras superfamily and directly activates mTORC1, promoting translation (*117, 118*). *PRKAG2* is a gamma subunit of AMPK (5′ AMP-activated protein kinase) known to activate TS (tuberous sclerosis) complex, which in turn inhibits RHEB function and therefore suppresses translational initiation (*117, 118*). Moreover, *PPARGC1A* is the most widely recognized

regulator for the metabolic adaptation of exercise (*119*). Although the preserved network was similarly activated by different exercise modes, resistance and endurance exercise should lead to distinct muscle adaptation, muscular hypertrophy and aerobic adaptation, respectively. Given the complex interplay of the positive and negative regulation of mTORC1 possibly triggered by *RHEB – PRKAG2 – PPARGC1A*, the preserved network may mediate differential translation orchestrating together with other translational regulators, including amino acid metabolism (*120*), TOP motif (*77*), or 5′ cap-dependent/-independent regulation mediated by mRNA secondary structure (*121, 122*). In this case, such different translational efficiency of each mRNA may become a driving force of exercise-mode-specific skeletal muscle adaptation.

To the best of my knowledge, this is the first study to investigate the effect of exercise on the gene co-expression network and its preservation. Using the notion of hub gene, I could suggest novel players that might play a pivotal role in mediating "open window" theory after acute endurance exercise in PBMC. Surprisingly, network preservation analysis revealed tightly connected networks preserved in different skeletal

muscle conditions. One preserved network, in which long-lasting network activation was observed in a time-course independent manner, can be associated with the repeated effect of acute resistance exercise, implying a key to answer the question why repeated exercise results in distinct muscle adaptation. The other preserved network was highly linked to mTORC1 and similarly activated by both resistance and endurance exercise, suggesting that this network-triggered translational regulation could further differentiate the distinct muscle adaptation to the different exercise modes. Given all, further studies are crucial to confirm the impact of the hub genes and preserved networks on exercise-induced systematic response and also the combined regulation of transcription and translation could be more focused to further elucidate the underlying mechanisms of muscle adaptation.

IV. CONCLUSIONS

In spite of the recent advances in genome-wide gene expression or proteomics analyses, there have been some drawbacks: 1. Translational regulation has been little focused and 2. Network-level dynamics of gene expression have never been investigated. Thus, the major aim of the current studies was to fill in the gaps in exercise/sports sciences and to enhance the understanding of what had little been focused using genome-wide approaches.

To tackle the first question, translational regulation, I utilized ribosome profiling to elucidate the followings: 1-a. Translationally regulated protein abundance, 1-b. Multiple protein species derived from a single mRNA (alternative translation), and 1-c. Alteration in translation speed. According to 1-a, I applied the original protocols of ribosome profiling to *in vivo* and successfully established *in vivo* ribosome profiling in mouse skeletal muscle for the first time. Depth-analysis of the ribosome profiling data in skeletal muscle revealed a striking discrepancy between transcriptional and translational profiles both before and after acute endurance exercise, indicative of a significant contribution of translational regulation in mouse skeletal muscle in a

transcription-level-independent manner. I could also suggest an orchestrated regulation of translation and protein degradation focusing on a specific gene target, in which transcriptional regulation had little effect.

As for 1-b, I developed an analytical pipeline to discover alternative translational events, including uORFs, N-terminal extension/truncation, and frameshifts. Many previously unrecognized alternative translational species have been newly identified in the current studies, one of which was associated with a deletion of its major functional domain, suggesting a potential impact on the regulation.

The novel biological notion discovered in 1-c is of particular interest. I developed a binary count method to examine at where and to what extent translating ribosomes can stall at a single-codon resolution both in *ex vivo* (RAW264 macrophages) and *in vivo* (mouse skeletal muscle). Although acute endurance exercise had little influence on translational speed, remarkably, LPS-induced acute inflammation significantly rendered translation speed (i.e., translational stalls) in RAW264 macrophages. I further confirmed

that such dynamics was clearly present in individual gene levels. More importantly, trypsin assay suggested that the altered translation speed induced conformational changes of the nascent protein. These outcomes imply a novel regulatory notion in biology, in which external stimuli-induced differential translation speed alters protein structure and function to maintain the cellular homeostasis. This is totally distinct from the conventional view, where external stimuli causes differential expression of mRNA or/and protein and changes the protein abundance to feed back to the physiological state.

To solve the second issue: 2. Network-level dynamics of gene expression have never been investigated, I used WGCNA and public database to capture and compare the network-level dynamics derived from different exercise modes, time-courses, and tissues. The network and hub gene analysis suggest potentially new players and the significance of the preserved gene expression network among different exercise modes/time-courses. Given that one of the preserved networks was associated with the key factors in translational regulation, gene co-expression network analysis may need to

be coupled with translational network to further understand differential muscle adaptation.

Overall, the current studies could address the major gaps and shed a novel light in exercise/sports sciences from a genome-wide perspective. However, much remain ambiguous. In particular, further researches are required to conceptually support the currently developed notion, to confirm the actual physiological impact, and for the details of the functional mechanisms.

ACKNOWLEDGEMENTS

Firstly, I would like to expression my sincere appreciation and gratitude to my supervisor Prof. Katsuhiko Suzuki for the continuous support of my Ph.D. study.

I would like to thank the rest of my thesis committee: Prof. Takashi Arao and Prof.

Shizuo Sakamoto for their insightful comments and advices.

I thank my labmates for all the fun we have had, which encouraged me to continue my research.

Last but not the least, I thank my families for so much support throughout my life. I would like to express my profound sense of appreciation to my friends: M.I., S.N., K.H., S.N., M.A., and Prof. Takashi for inspiring me and paving the way for my Ph.D. life when I was still in Brisbane. Lastly, I would like to say “thank you” from the bottom of my heart. Without you, Y.N., I might have not come back to Japan and never achieved the current work. Thank you.

REFERENCES

1. H. Sako, K. Suzuki, Exploring the importance of translational regulation in the inflammatory responses by a genome-wide approach. *Exerc Immunol Rev* **20**, 55-67 (2014).

2. K. Ohlendieck, Skeletal muscle proteomics: current approaches, technical challenges and emerging techniques. *Skelet Muscle* **1**, 6 (2011).

3. B. Wittmann-Liebold, H. R. Graack, T. Pohl, Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. *Proteomics* **6**, 4688-4703 (2006).

4. J. A. Paulo, V. Kadiyala, P. A. Banks, H. Steen, D. L. Conwell, Mass spectrometry-based proteomics for translational research: a technical overview. *Yale J Biol Med* **85**, 59-73 (2012).

5. R. J. Dobbs, A. Charlett, A. G. Purkiss, S. M. Dobbs, C. Weller, D. W. Peterson, Association of circulating TNF-alpha and IL-6 with ageing and parkinsonism. *Acta Neurol Scand* **100**, 34-41 (1999).

6. M. Bullo, P. Garcia-Lorda, I. Megias, J. Salas-Salvado, Systemic inflammation, adipose tissue tumor necrosis factor, and leptin expression. *Obes Res* **11**, 525-531 (2003).

7. M. H. Elliott, D. S. Smith, C. E. Parker, C. Borchers, Current trends in quantitative proteomics. *J Mass Spectrom* **44**, 1637-1660 (2009).

8. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009).

9. M. P. Washburn, A. Koller, G. Oshiro, R. R. Ulaszek, D. Plouffe, C. Deciu, E. Winzeler, J. R. Yates, 3rd, Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**, 3107-3112 (2003).

10. Q. Tian, S. B. Stepaniants, M. Mao, L. Weng, M. C. Feetham, M. J. Doyle, E. C. Yi, H. Dai, V. Thorsson, J. Eng, D. Goodlett, J. P. Berger, B. Gunter, P. S. Linseley, R. B. Stoughton, R. Aebersold, S. J. Collins, W. A. Hanlon, L. E. Hood, Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics* **3**, 960-969 (2004).

11. L. Nie, G. Wu, W. Zhang, Correlation between mRNA and protein abundance in Desulfovibrio vulgaris: a multiple regression to identify sources of variations. *Biochem Biophys Res Commun* **339**, 603-610 (2006).

12. S. P. Gygi, Y. Rochon, B. R. Franza, R. Aebersold, Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730 (1999).

13. D. Greenbaum, C. Colangelo, K. Williams, M. Gerstein, Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4**, 117 (2003).

14. B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

15. D. Chinkes, S. Klein, X. J. Zhang, R. R. Wolfe, Infusion of labeled KIC is more accurate than labeled leucine to determine human muscle protein synthesis. *Am J Physiol* **270**, E67-71 (1996).

16. E. I. Glover, S. M. Phillips, B. R. Oates, J. E. Tang, M. A. Tarnopolsky, A. Selby, K. Smith, M. J. Rennie, Immobilization induces anabolic resistance in human myofibrillar protein synthesis with low and high dose amino acid infusion. *J Physiol* **586**, 6049-6061 (2008).

17. S. M. Phillips, K. D. Tipton, A. Aarsland, S. E. Wolf, R. R. Wolfe, Mixed muscle protein synthesis and breakdown after resistance exercise in humans. *Am J Physiol* **273**, E99-107 (1997).

18. M. J. Rennie, R. H. Edwards, D. Halliday, D. E. Matthews, S. L. Wolman, D. J. Millward, Muscle protein synthesis measured by stable isotope techniques in man: the effects of feeding and fasting. *Clin Sci (Lond)* **63**, 519-523 (1982).

19. J. J. Fyfe, D. J. Bishop, N. K. Stepto, Interference between concurrent resistance and endurance exercise: molecular bases and the role of individual training variables. *Sports Med* **44**, 743-762 (2014).

20. V. Kumar, P. Atherton, K. Smith, M. J. Rennie, Human muscle protein synthesis and breakdown during and after exercise. *J Appl Physiol*

(1985) **106**, 2026-2039 (2009).

21. R. Ogasawara, K. Sato, K. Matsutani, K. Nakazato, S. Fujita, The order of concurrent endurance and resistance exercise modifies mTOR signaling and protein synthesis in rat skeletal muscle. *Am J Physiol Endocrinol Metab* **306**, E1155-1162 (2014).

22. D. L. Williamson, N. Kubica, S. R. Kimball, L. S. Jefferson, Exercise-induced alterations in extracellular signal-regulated kinase 1/2 and mammalian target of rapamycin (mTOR) signalling to regulatory mechanisms of mRNA translation in mouse muscle. *J Physiol* **573**, 497-510 (2006).

23. N. T. Ingolia, G. A. Brar, S. Rouskin, A. M. McGeachy, J. S. Weissman, The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**, 1534-1550 (2012).

24. J. Medenbach, M. Seiler, M. W. Hentze, Translational control via protein-regulated upstream open reading frames. *Cell* **145**, 902-913

(2011).

25. R. J. Jackson, C. U. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127 (2010).

26. S. Chatterjee, J. K. Pal, Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell* **101**, 251-262 (2009).

27. N. Sonenberg, A. G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745 (2009).

28. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802 (2011).

29. W. Zhou, W. Song, Leaky scanning and reinitiation regulate BACE1 gene expression. *Mol Cell Biol* **26**, 3353-3364 (2006).

30. F. T. Lin, O. A. MacDougald, A. M. Diehl, M. D. Lane, A 30-kDa alternative translation product of the CCAAT/enhancer binding

protein alpha message: transcriptional activator lacking antimitotic activity. *Proc Natl Acad Sci U S A* **90**, 9606-9610 (1993).

31. S. Matsufuji, T. Matsufuji, Y. Miyazaki, Y. Murakami, J. F. Atkins, R. F. Gesteland, S. Hayashi, Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80**, 51-60 (1995).

32. A. E. Pegg, Recent advances in the biochemistry of polyamines in eukaryotes. *Biochem J* **234**, 249-262 (1986).

33. M. Zhou, J. Guo, J. Cha, M. Chae, S. Chen, J. M. Barral, M. S. Sachs, Y. Liu, Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**, 111-115 (2013).

34. Y. Xu, P. Ma, P. Shah, A. Rokas, Y. Liu, C. H. Johnson, Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* **495**, 116-120 (2013).

35. B. Seidelt, C. A. Innis, D. N. Wilson, M. Gartmann, J. P. Armache, E. Villa, L. G. Trabuco, T. Becker, T. Mielke, K. Schulten, T. A. Steitz, R. Beckmann, Structural insight into nascent polypeptide

chain-mediated translational stalling. *Science* **326**, 1412-1415 (2009).

36. S. Arenz, H. Ramu, P. Gupta, O. Berninghausen, R. Beckmann, N. Vazquez-Laslop, A. S. Mankin, D. N. Wilson, Molecular basis for erythromycin-dependent ribosome stalling during translation of the ErmBL leader peptide. *Nat Commun* **5**, 3501 (2014).

37. S. Arenz, S. Meydan, A. L. Starosta, O. Berninghausen, R. Beckmann, N. Vazquez-Laslop, D. N. Wilson, Drug sensing by the ribosome induces translational arrest via active site perturbation. *Mol Cell* **56**, 446-452 (2014).

38. S. Bhushan, H. Meyer, A. L. Starosta, T. Becker, T. Mielke, O. Berninghausen, M. Sattler, D. N. Wilson, R. Beckmann, Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* **40**, 138-146 (2010).

39. T. Tuller, Y. Y. Waldman, M. Kupiec, E. Ruppin, Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A* **107**, 3645-3650 (2010).

40. S. Mahlab, M. Linial, Speed controls in translating secretory proteins in eukaryotes--an evolutionary perspective. *PLoS Comput Biol* **10**, e1003294 (2014).

41. J. Gardin, R. Yeasmin, A. Yurovsky, Y. Cai, S. Skiena, B. Futcher, Measurement of average decoding rates of the 61 sense codons in vivo. *Elife* **3**, (2014).

42. A. T. Martens, J. Taylor, V. J. Hilser, Ribosome A and P sites revealed by length analysis of ribosome profiling data. *Nucleic Acids Res* 10.1093/nar/gkv200, (2015).

43. J. Lu, W. R. Kobertz, C. Deutsch, Mapping the electrostatic potential within the ribosomal exit tunnel. *J Mol Biol* **371**, 1378-1391 (2007).

44. J. Lu, C. Deutsch, Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol* **384**, 73-86 (2008).

45. C. A. Charneski, L. D. Hurst, Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol* **11**, e1001508 (2013).

46. A. M. Michel, K. R. Choudhury, A. E. Firth, N. T. Ingolia, J. F. Atkins, P. V. Baranov, Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* **22**, 2219-2229 (2012).

47. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559 (2008).

48. A. Neueder, G. P. Bates, A common gene expression signature in Huntington's disease patient brain regions. *BMC Med Genomics* **7**, 60 (2014).

49. A. P. Gasch, M. B. Eisen, Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol* **3**, RESEARCH0059 (2002).

50. A. Ghazalpour, S. Doss, B. Zhang, S. Wang, C. Plaisier, R. Castellanos, A. Brozell, E. E. Schadt, T. A. Drake, A. J. Lusis, S. Horvath, Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet* **2**, e130 (2006).

51. S. Horvath, B. Zhang, M. Carlson, K. V. Lu, S. Zhu, R. M. Felciano, M. F. Laurance, W. Zhao, S. Qi, Z. Chen, Y. Lee, A. C. Scheck, L. M. Liau, H. Wu, D. H. Geschwind, P. G. Febbo, H. I. Kornblum, T. F. Cloughesy, S. F. Nelson, P. S. Mischel, Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A* **103**, 17402-17407 (2006).

52. Y. Huang, H. Li, H. Hu, X. Yan, M. S. Waterman, H. Huang, X. J. Zhou, Systematic discovery of functional modules and context-specific functional annotation of human genome. *Bioinformatics* **23**, i222-229 (2007).

53. M. P. Keller, Y. Choi, P. Wang, D. B. Davis, M. E. Rabaglia, A. T. Oler, D. S. Stapleton, C. Argmann, K. L. Schueler, S. Edwards, H. A. Steinberg, E. Chaibub Neto, R. Kleinhanz, S. Turner, M. K. Hellerstein, E. E. Schadt, B. S. Yandell, C. Kendziorski, A. D. Attie, A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Res* **18**,

706-716 (2008).

54. M. J. Mason, G. Fan, K. Plath, Q. Zhou, S. Horvath, Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* **10**, 327 (2009).

55. C. Niehrs, N. Pollet, Synexpression groups in eukaryotes. *Nature* **402**, 483-487 (1999).

56. J. M. Stuart, E. Segal, D. Koller, S. K. Kim, A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249-255 (2003).

57. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005).

58. J. Jiang, P. Jia, Z. Zhao, B. Shen, Key regulators in prostate cancer identified by co-expression module analysis. *BMC Genomics* **15**, 1015 (2014).

59. J. Albright, P. M. Quizon, A. J. Lusis, B. J. Bennett, Genetic network

identifies novel pathways contributing to atherosclerosis susceptibility in the innominate artery. *BMC Med Genomics* **7**, 51 (2014).

60. P. Langfelder, P. S. Mischel, S. Horvath, When is hub gene selection better than standard meta-analysis? *PLoS One* **8**, e61505 (2013).

61. A. Kasprzyk, BioMart: driving a paradigm change in biological data management. *Database (Oxford)* **2011**, bar049 (2011).

62. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

63. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* **11**, R25 (2010).

64. D. J. Lynn, G. L. Winsor, C. Chan, N. Richard, M. R. Laird, A. Barsky, J. L. Gardy, F. M. Roche, T. H. Chan, N. Shah, R. Lo, M. Naseer, J. Que, M. Yau, M. Acab, D. Tulpan, M. D. Whiteside, A. Chikatamarla, B.

Mah, T. Munzner, K. Hokamp, R. E. Hancock, F. S. Brinkman, InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Molecular systems biology* **4**, 218 (2008).

65. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

66. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).

67. H. Guo, N. T. Ingolia, J. S. Weissman, D. P. Bartel, Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).

68. E. P. O'Brien, S. T. Hsu, J. Christodoulou, M. Vendruscolo, C. M. Dobson, Transient tertiary structure formation within the ribosome exit port. *J Am Chem Soc* **132**, 16928-16937 (2010).

69. L. D. Cabrita, C. M. Dobson, J. Christodoulou, Protein folding on the

ribosome. *Curr Opin Struct Biol* **20**, 33-45 (2010).

70. M. N. McCall, B. M. Bolstad, R. A. Irizarry, Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242-253 (2010).

71. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).

72. B. H. Zhang, S., A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4**, Article 17 (2005).

73. M. A. Freeberg, T. Han, J. J. Moresco, A. Kong, Y. C. Yang, Z. J. Lu, J. R. Yates, J. K. Kim, Pervasive and dynamic protein binding sites of the mRNA transcriptome in Saccharomyces cerevisiae. *Genome Biol* **14**, R13 (2013).

74. F. Palmieri, The mitochondrial transporter family (SLC25): physiological and pathological implications. *Pflugers Arch* **447**, 689-709 (2004).

75. E. R. Kunji, The role and structure of mitochondrial carriers. *FEBS Lett* **564**, 239-244 (2004).

76. R. P. Anunciado-Koza, J. Zhang, J. Ukropec, S. Bajpeyi, R. A. Koza, R. C. Rogers, W. T. Cefalu, R. L. Mynatt, L. P. Kozak, Inactivation of the mitochondrial carrier SLC25A25 (ATP-Mg2+/Pi transporter) reduces physical endurance and metabolic efficiency in mice. *J Biol Chem* **286**, 11659-11671 (2011).

77. C. C. Thoreen, L. Chantranupong, H. R. Keys, T. Wang, N. S. Gray, D. M. Sabatini, A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109-113 (2012).

78. T. A. Gautsch, J. C. Anthony, S. R. Kimball, G. L. Paul, D. K. Layman, L. S. Jefferson, Availability of eIF4E regulates skeletal muscle protein synthesis during recovery from exercise. *Am J Physiol* **274**, C406-414 (1998).

79. W. Wang, C. J. Schulze, W. L. Suarez-Pinzon, J. R. Dyck, G. Sawicki, R. Schulz, Intracellular action of matrix metalloproteinase-2 accounts for

acute myocardial ischemia and reperfusion injury. *Circulation* **106**, 1543-1549 (2002).

80. T. J. Hohn, T. Grune, The proteasome and the degradation of oxidized proteins: part III-Redox regulation of the proteasomal system. *Redox Biol* **2**, 388-394 (2014).

81. M. Koppen, T. Langer, Protein degradation within mitochondria: versatile activities of AAA proteases and other peptidases. *Critical reviews in biochemistry and molecular biology* **42**, 221-242 (2007).

82. A. M. Pickering, R. A. Linder, H. Zhang, H. J. Forman, K. J. Davies, Nrf2-dependent induction of proteasome and Pa28alphabeta regulator are required for adaptation to oxidative stress. *J Biol Chem* **287**, 10021-10031 (2012).

83. R. K. Jobling, M. Assoum, O. Gakh, S. Blaser, J. A. Raiman, C. Mignot, E. Roze, A. Durr, A. Brice, N. Levy, C. Prasad, T. Paton, A. D. Paterson, N. M. Roslin, C. R. Marshall, J. P. Desvignes, N. Roeckel-Trevisiol, S. W. Scherer, G. A. Rouleau, A. Megarbane, G. Isaya, V. Delague, G.

Yoon, PMPCA mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. *Brain : a journal of neurology* **138**, 1505-1517 (2015).

84. O. Gakh, P. Cavadini, G. Isaya, Mitochondrial processing peptidases. *Biochimica et biophysica acta* **1592**, 63-77 (2002).

85. B. Lu, C. Poirier, T. Gaspar, C. Gratzke, W. Harrison, D. Busija, M. M. Matzuk, K. E. Andersson, P. A. Overbeek, C. E. Bishop, A mutation in the inner mitochondrial membrane peptidase 2-like gene (Immp2l) affects mitochondrial function and impairs fertility in mice. *Biology of reproduction* **78**, 601-610 (2008).

86. C. Barbosa, I. Peixeiro, L. Romao, Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet* **9**, e1003529 (2013).

87. D. R. Morris, A. P. Geballe, Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**, 8635-8642 (2000).

88. E. Arana, A. Vehlow, N. E. Harwood, E. Vigorito, R. Henderson, M.

Turner, V. L. Tybulewicz, F. D. Batista, Activation of the small GTPase Rac2 via the B cell receptor regulates B cell adhesion and immunological-synapse formation. *Immunity* **28**, 88-99 (2008).

89. F. Guo, J. A. Cancelas, D. Hildeman, D. A. Williams, Y. Zheng, Rac GTPase isoforms Rac1 and Rac2 play a redundant and crucial role in T-cell development. *Blood* **112**, 1767-1775 (2008).

90. A. Sengupta, J. Arnett, S. Dunn, D. A. Williams, J. A. Cancelas, Rac2 GTPase deficiency depletes BCR-ABL+ leukemic stem cells and progenitors in vivo. *Blood* **116**, 81-84 (2010).

91. M. J. Walmsley, S. K. Ooi, L. F. Reynolds, S. H. Smith, S. Ruf, A. Mathiot, L. Vanes, D. A. Williams, M. P. Cancro, V. L. Tybulewicz, Critical roles for Rac1 and Rac2 GTPases in B cell development and signaling. *Science* **302**, 459-462 (2003).

92. M. Kozak, Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13-37 (2005).

93. E. G. Magny, J. I. Pueyo, F. M. Pearl, M. A. Cespedes, J. E. Niven, S. A.

Bishop, J. P. Couso, Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116-1120 (2013).

94. C. H. Yu, Y. Dang, Z. Zhou, C. Wu, F. Zhao, M. S. Sachs, Y. Liu, Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular cell* **59**, 744-754 (2015).

95. P. B. O'Connor, G. W. Li, J. S. Weissman, J. F. Atkins, P. V. Baranov, rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. *Bioinformatics* **29**, 1488-1491 (2013).

96. G. W. Li, E. Oh, J. S. Weissman, The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538-541 (2012).

97. J. P. Luzio, P. R. Pryor, N. A. Bright, Lysosomes: fusion and function. *Nat Rev Mol Cell Biol* **8**, 622-632 (2007).

98. C. K. Ferreira, J. Prestes, F. F. Donatto, R. Verlengia, J. W. Navalta, C.

R. Cavaglieri, Phagocytic responses of peritoneal macrophages and neutrophils are different in rats following prolonged exercise. *Clinics (Sao Paulo)* **65**, 1167-1173 (2010).

99. R. V. dos Santos, E. C. Caperuto, M. T. de Mello, L. F. Costa Rosa, Effect of exercise on glutamine metabolism in macrophages of trained rats. *Eur J Appl Physiol* **107**, 309-315 (2009).

100. B. K. Pedersen, H. Bruunsgaard, How physical exercise influences the establishment of infections. *Sports Med* **19**, 393-400 (1995).

101. M. W. Kakanis, J. Peake, E. W. Brenu, M. Simmonds, B. Gray, S. L. Hooper, S. M. Marshall-Gradisnik, The open window of susceptibility to infection after acute exercise in healthy young male elite athletes. *Exerc Immunol Rev* **16**, 119-137 (2010).

102. S. J. Szabo, B. M. Sullivan, S. L. Peng, L. H. Glimcher, Molecular mechanisms regulating Th1 immune responses. *Annu Rev Immunol* **21**, 713-758 (2003).

103. G. Miklossy, T. S. Hilliard, J. Turkson, Therapeutic modulators of

STAT signalling for human diseases. *Nat Rev Drug Discov* **12**, 611-629 (2013).

104. T. Decker, M. Muller, S. Stockinger, The yin and yang of type I interferon activity in bacterial infection. *Nat Rev Immunol* **5**, 675-687 (2005).

105. K. S. Wang, J. Ritz, D. A. Frank, IL-2 induces STAT4 activation in primary NK cells and NK cell lines, but not in T cells. *J Immunol* **162**, 299-304 (1999).

106. M. Hatakeyama, M. Tsudo, S. Minamoto, T. Kono, T. Doi, T. Miyata, M. Miyasaka, T. Taniguchi, Interleukin-2 receptor beta chain gene: generation of three receptor forms by cloned human alpha and beta chain cDNA's. *Science* **244**, 551-556 (1989).

107. Y. Takata, D. Hamada, K. Miyatake, S. Nakano, F. Shinomiya, C. R. Scafe, V. M. Reeve, D. Osabe, M. Moritani, K. Kunika, N. Kamatani, H. Inoue, N. Yasui, M. Itakura, Genetic association between the PRKCH gene encoding protein kinase Ceta isozyme and rheumatoid arthritis

in the Japanese population. *Arthritis Rheum* **56**, 30-42 (2007).

108. Y. M. Peng, M. D. van de Garde, K. F. Cheng, P. A. Baars, E. B. Remmerswaal, R. A. van Lier, C. R. Mackay, H. H. Lin, J. Hamann, Specific expression of GPR56 by human cytotoxic lymphocytes. *J Leukoc Biol* **90**, 735-740 (2011).

109. S. Shashidhar, G. Lorente, U. Nagavarapu, A. Nelson, J. Kuo, J. Cummins, K. Nikolich, R. Urfer, E. D. Foehr, GPR56 is a GPCR that is overexpressed in gliomas and functions in tumor cell adhesion. *Oncogene* **24**, 1673-1682 (2005).

110. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

111. C. Grandori, N. Gomez-Roman, Z. A. Felton-Edkins, C. Ngouenet, D. A. Galloway, R. N. Eisenman, R. J. White, c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nat Cell Biol* **7**, 311-318 (2005).

112. I. Schlosser, M. Holzel, M. Murnseer, H. Burtscher, U. H. Weidle, D. Eick, A role for c-Myc in the regulation of ribosomal RNA processing. *Nucleic Acids Res* **31**, 6148-6156 (2003).

113. A. Gandarillas, F. M. Watt, c-Myc promotes differentiation of human epidermal stem cells. *Genes Dev* **11**, 2869-2882 (1997).

114. A. Wilson, M. J. Murphy, T. Oskarsson, K. Kaloulis, M. D. Bettess, G. M. Oser, A. C. Pasche, C. Knabenhans, H. R. Macdonald, A. Trumpp, c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* **18**, 2747-2763 (2004).

115. B. Hoffman, D. A. Liebermann, Apoptotic signaling by c-MYC. *Oncogene* **27**, 6462-6472 (2008).

116. A. J. Murton, R. Billeter, F. B. Stephens, S. G. Des Etages, F. Graber, R. J. Hill, K. Marimuthu, P. L. Greenhaff, Transient transcriptional events in human skeletal muscle at the outset of concentric resistance exercise training. *J Appl Physiol (1985)* **116**, 113-125 (2014).

117. X. M. Ma, J. Blenis, Molecular mechanisms of mTOR-mediated

translational control. *Nat Rev Mol Cell Biol* **10**, 307-318 (2009).

118. C. C. Dibble, B. D. Manning, Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. *Nat Cell Biol* **15**, 555-564 (2013).

119. M. C. Chan, Z. Arany, The many roles of PGC-1alpha in muscle--recent developments. *Metabolism* **63**, 441-451 (2014).

120. S. G. Kim, G. R. Hoffman, G. Poulogiannis, G. R. Buel, Y. J. Jang, K. W. Lee, B. Y. Kim, R. L. Erikson, L. C. Cantley, A. Y. Choo, J. Blenis, Metabolic stress controls mTORC1 lysosomal localization and dimerization by regulating the TTT-RUVBL1/2 complex. *Mol Cell* **49**, 172-185 (2013).

121. G. C. Scheper, M. S. van der Knaap, C. G. Proud, Translation matters: protein synthesis defects in inherited disease. *Nat Rev Genet* **8**, 711-723 (2007).

122. A. S. Lee, P. J. Kranzusch, J. H. Cate, eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*

522, 111-114 (2015).

FIGURE LEGENDS

Figure 1. Schematic representation of modified ribosome profiling.

Figure 2. Reproducibility of mRNA-Seq and ribosome profiling in RAW264.

Pearson correlations of biological replicates (log_2 scale aligned reads) are shown.

Figure 3. Triplet periodicity of 5′ read positions in mRNA-Seq and ribosome profiling in RAW264.

Main coding frame is indicated by “0”.

Figure 4. Reproducibility of mRNA-Seq and ribosome profiling in skeletal muscle.

Pearson correlations of biological replicates (log_2 scale aligned reads) are shown.

Figure 5. Triplet periodicity of 5′ read positions in mRNA-Seq and ribosome profiling in skeletal muscle.

Main coding frame is indicated by “0”.

Figure 6. Determination of minimum read counts for downstream analyses.

For reliable downstream measurements, minimum read counts were determined using the independent biological replicates of ribosome profiling (immediately after exercise). For each gene, the fraction of reads (read counts of one replicate divided by total read counts derived from both replicates) was calculated. Based on average read counts,

genes were binned and standard deviation (SD) of the fraction was computed for each bin. The SD for counting statistics between the replicates was predicted using binominal partitioning of the read counts. The minimum read count (125 RPM) is indicated by a dashed line from where error between biological replicates becomes stably larger than error from counting statistics. RPM: Reads Per Million.

Figure 7. Principal component analysis of mRNA-Seq and ribosome profiling.

Principal component analysis was conducted for each replicate targeting genes with more than 125 RPM (n = 1011). The replicates belonging to the same condition were circle together. The replicate derived from mRNA-Seq or ribosome profiling was surrounded by a rectangle colored in grey or black, respectively.

Figure 8. Transcriptional or translational changes with or without exercise.

Log-transformed expression ratio was compared using MA plot targeting genes with more than 125 RPM (n = 1011). Differentially expressed genes ($P < 0.01$) were determined by edgeR and colored in red.

Figure 9. Quantification of transcription and translation levels of Slc25a25.

The relative abundance of Slc25a25 transcription and translation, measured by mRNA-Seq and ribosome profiling, respectively, were compared with and without exercise using ribosome profiling. Mean ± SD.

Figure 10. Exercise-induced change in translational efficiency of TOP- and non TOP motif genes.

Log-transformed translational efficiencies (RPF/mRNA-Seq) and the ratios (exercise/no exercise) were calculated for known TOP motif genes (n = 59) and the other non TOP motif genes (n = 953). Kolmogorov-Smirnov test was used for statistics.

Figure 11. Quantification of transcription and protein levels of Slc25a25.

(A) Slc25a25 protein abundance was analyzed by western blotting (n = 6). Mean ± SE. One-way ANOVA with Tukey (* $P < 0.01$). (B) Slc25a25 mRNA level was measured by real time PCR (n = 6). Mean ± SE. One-way ANOVA with Tukey (* $P < 0.01$).

Figure 12. Measurements of the expression levels of proteolysis related genes.

Genes related to proteolysis were measured by real time qPCR (n = 6). Mean ± SE. One-way ANOVA with Tukey (** $P < 0.01$, * $P < 0.05$).

Figure 13. Correlation and multiple regression analyses.

Association strength was measured among the protein abundance of Slc25a25 and gene expressions of Slc25a25, Mmp2, Nrf2, Atrogin1, MuRF1, Immp2l, Yme1l1, and Pmpca using simple correlation analysis and multiple linear regression analysis. In the table of the regression analysis, an estimated coefficient of each gene and a corresponding *P* value are shown.

Figure 14. Confirmation of previously known uORF and alternative translational start site.

Previously recognized AUG and non-AUG start sites of uORFs and N-terminal extension are shown.

Figure 15. Previously unknown alternative translation in Alcohol dehydrogenase.

Main coding frame of the gene is indicated by “frame 0” (blue plots, blue arrow, blue bar) and alternative translation is shown in green (green plots, green arrow, green bar), “frame 1”. Start and stop codons are shown in blue and red lines. “Upstream” and “Downstream” indicate the mRNA regions upstream and downstream of the overlapped region of the alternative translation and the canonical translation.

Figure 16. N-terminal truncation in RAC2.

Novel N-terminal truncation in RAC2 was identified and validated by western blotting. Main coding frame of the gene is indicated by “frame 0” (blue plots, blue arrow, blue bar). Truncated region is shown in gashed blue arrow in LPS 0min. Red arrow indicates another alternative translation. “Upstream” and “Downstream” indicate the mRNA regions upstream and downstream of the overlapped region of the alternative translation and the canonical translation. DEX: Dexamethasone.

Figure 17. Rac2 alternative start codon preservation across species.

Annotated and alternative start codon preservation is shown. Franking sequences of annotated start codon are also shown for Koazk sequence comparison.

Figure 18. Novel translational isoform in skeletal muscle, Ldha.

Figure 19. Novel translational isoform in skeletal muscle, Tip1.

Figure 20. Validation of the binary approach.

(A) Comparison of ribosome density patterns between the raw and binary counts. Top: accumulated raw RPF counts and positions of all the reference genes shown in black. Bottom: accumulated binary RPF counts and positions of all the reference genes shown in orange. Magnified views (-20 to 20 nt and 200 to 300 nt regions) were also shown.

(B) Cumulated coefficient variants (CV) for the raw (black) and binary counts (red). Using all the reference genes, CV of the ribosome density of the main coding frame were calculated from the start codons of all the genes to the last codons.

Kolmogorov-Smirnov test was used for the statistics.

Figure 21. Schematic image of the measurement of translation speed.

Screening windows are shown in which there is a residue of interest flanked by 10 and 24 codons. Example RPF patterns are shown in red, derived from the accumulated binary counts at each identical relative position in different screening windows. Pattern 1: In the screening window, if codons (residue) of interest cause translational arrest when the target residue codon is about to enter the ribosome, a density peak would be observed as indicated by the peak 1 (i.e., 10 codons away from the target codon). Pattern 2: If a residue of interest causes translational arrest when the residue is near the end of ribosomal exit tunnel, a density peak would be observed as indicated by the peak 2. RPF: ribosome-protected mRNA fragments.

Figure 22. Translational stalls in TAC sites in mouse.

Superimposed 20 ribosome density patterns in RAW264 macrophages. (top) Without binary approach or (bottom) with binary approach, 20 different ribosome density patters,

each of which reflects the effect of single residue of interest on translation speed in RAW264, are superimposed. The A site in ribosomal complex is shown in position 0 ~ +2 nt. Position +3 ~ +5 nt indicates the P site. Translational arrest core (TAC) sites, from A site to positions 5 residues away from P-site, are indicated by two dashed grey lines. Ribosome density at each base position was normalized by the mRNA abundance at the corresponding base position.

Figure 23-29. Dynamics in translational stalls in basal and LPS-stimulated conditions in RAW264.

Translational stall derived from each residue was characterized in basal and LPS-induced inflammatory states. * $P < 0.05$, Smirnov-Grubbs test.

Figure 30, 31. Reproducibility of translational stall measurements (61 codons).

Reproducibility of ribosome densities, induced by each of 61 codons, was confirmed for both A/P sites and the other TAC sites under the basal and LPS-stimulated conditions.

Figure 32. Alteration of translational stall induced by 61 codons in acute inflammatory and acute endurance exercise conditions.

Bi-plots exhibit the effect of external stimulus, LPS (A) and exercise (B), on translational stall induced by each of 61 codons. Each dot represents a set of a codon and the corresponding position in ribosomal complex, at which the codon affects translation. Colored dots show differentially regulated stalls in RAW264 ($P < 0.05$ and Z score difference > 1.5) and in skeletal muscle ($P < 0.05$ and Z score difference > 1.2).

Figure 33. Conditional difference in GAC-associated read length.

Read length in RPF of ribosome profiling or mRNA of mRNA-Seq is shown when GAC codon is located at an indicated position (Ex5, Ex4, or Ex3) in the basal (blue) or LPS-stimulated condition (red).

Figure 34. LPS-induced decrease in translational stall at GAC codon site in Calreticulin.

The position of Asp codons are shown in grey lines and the GAC codon indicated by arrows.

Figure 35. Trypsin assay for Calreticulin and GAPDH.

Total protein lysate was trypsin-digested for 30 min at indicated trypsin concentrations. Intact protein of interest after trypsin digestion was divided by control group (without trypsin) to obtain relative protein level (Intact/No degradation). Blue: Basal, Red: LPS 120min. In Calreticulin, bacasue digested Calreticulin was observed, the ratio (Intact/Degraded) was also calculated. One-way ANOVA with TukeyHSD. Mean ± SE, n = 6 except n = 5 at 20 ng/μl trypsin (Intact/No degradation) where 1 outlier was removed by Grubs test.

Figure 36. Current research design for gene co-expression network preservation analysis.

The top describes the comparison of PBMC network preservation across tissues: (I) preservation between PBMC and VL in acute endurance exercise (post 2.5 hr and 5 hr), (II) preservation between PBMC and VL in acute resistance exercise (post 4 hr), and (III) preservation between PBMC and VL in acute resistance exercise (post 24 hr).

Preservation in all (I), (II), and (III) indicates tissue-independent preservation. The bottom describes the comparison of VL in acute resistance (post 4 hr) network with the other conditions: (1) preservation between VL in acute endurance (post 4 hr) and PBMC, (2) preservation between VL in acute resistance (post 4 hr) and VL in acute endurance (post 2.5 hr and 5 hr), and (3) preservation between VL in acute endurance (post 4 hr) and VL in acute resistance (post 24 hr). Preservation in (2) and (3) indicates muscle-specific preservation. Preservation only in (2) suggests the network is present specifically early stages after exercise in an exercise-mode-independent manner. Preservation only in (3) suggests the network is resistance-exercise-dependent but time-course-independent. PBMC: Peripheral Blood Mononuclear Cell, VL: Vastus Lateralis.

Figure 37. Cluster dendrogram construction.

Cluster dendrogram based on the calculated adjacency matrix of PBMC gene co-expression network. To identify the network modules, a dynamic hybrid tree cutting algorithm was used and then highly co-expressed modules ($r > 0.90$) were merged.

Different module colors indicate different network modules.

Figure 38. Detection of significantly correlated network in PBMC to exercise trait.

Correlation between eigengene and exercise traits. To identify biologically interesting modules, Pearson correlations between eigengene (the representative or first principal component of the module) and exercise traits were examined. Exercise: comparison between pre- and post-exercise (both 0 min and 1 hr were included). ExercisePre0min: comparison between pre- and immediate after exercise. ExercisePre1h: comparison between pre- and post 1 hr of exercise. ExerciseTime: comparison between immediate after exercise and post 1hr of exercise. Batch: comparison between two different datasets we combined for PBMC network analysis. ME: Module Eigengene. Different color names refer different network modules.

Figure 39. Module correlation in PBMC.

Correlation between module membership and absolute gene significance for each exercise trait (ExercisePre1hr). Module membership refers the strength of the

connection between each module member gene and the eigengene in the module. Absolute gene significance is the absolute correlation between gene expression patters and an exercise trait.

Figure 40. Visualization of module network in PBMC.

Network visualization for the top 1000 strongest network edges of each selected module. Five modules (blue, green, magenta, red, and turquoise) were selected based on the correlation strength between module eigengene/module membership and pathway enrichment. The network was visualized by Cytoscape and the node (gene) positions were calculated by edge-weighted spring embedded method. Edge weight is the connection strength. A stronger connection is illustrated by a thicker, darker, and shorter edge. Node color and size indicate a different module and summed connection strength of the node to other nodes, respectively.

Figure 41-43. Preservation of module network of PBMC in VL with different conditions.

Module preservation was measured by $Z_{summary}$ and medianRank for the selected modules in PBMC network (Figure 41). $Z_{summary}$ with less than 2 (shown in red line) indicates no preservation. $Z_{summary}$ is the summary statistics of $Z_{density}$ (Figure 42) and $Z_{connectivity}$ (Figure 43). Because $Z_{summary}$ often depends on module size, relative preservation statistics (medianRank) was also calculated.

Figure 44. Schematic images of hub status preservation.

(A) Schematic representation of module/hub preservation focusing on the module in the reference network. In Figure 43, module-based preservation analysis was conducted, measuring how strongly the connectivity was preserved between the module members and the hub genes. Top: a module in a reference module, Bottom: preservation status of the reference module in test network. Solid lines and dashed lines indicate preserved/strong and not preserved/weaken connections, respectively. H: hub gene.

(B) Schematic representation of hub status preservation irrespective of the reference module members. Differently from the analysis in Figure 43, hub status-based

preservation was analyzed to measure to what extent hub characteristics (high kME) remains in a test network. Top: a hub in a module of a reference module, Bottom: preservation status of the hub gene in test network. Bottom left: Even if the connectivity between the hub and the reference module members becomes weak in the test network, other module members can connect strongly to the hub, enabling the hub still strong hub in a new module of the test network (i.e., high kME values derived from the green module). Bottom right: In contrast, even if the connectivity between the hub and the reference module members remains strong in the test network, another hub (H′) in the same module has stronger impact on the new module, altering the hub (H) to a relatively weaker hub in the test network. H′: another hub gene.

Figure 45. The kME and GS of hub genes in the most preserved module of PBMC (turquoise) colored in turquoise and the least preserved module (magenta) colored in magenta. Grey dots indicate the least hub genes in PBMC network irrespective of any module. In the turquoise module, genes with kME > 0.8 and GS > 0.5 were first

selected. In the magenta module which is relatively smaller than turquoise, genes with kME > 0.7 and GS > 0.5 were first selected. Top 20 and top 13 hub genes (all the genes that passed the criteria in the magenta module) in the turquoise and magenta module were selected for analysis, respectively. The hub genes were ranked by the order of the absolute values of kME multiplied by absolute GS. The least hub genes shown in grey were the genes with kME < 0.3 and GS < 0.1 in any module in PBMC. GS values were based on the comparison between pre and post 1 hr in PBMC. kME: eigengene-based connectivity, GS: gene significance.

Figure 46,47. Status of the PBMC hub genes and PBMC least hub genes in skeletal muscle conditions (acute resistance 24 hr, acute resistance 4 hr, and acute endurance). Differently from the preservation analysis $Z_{connection}$ focusing on the connectivity to hub genes among the module members, here focused on the highest kME the selected hub gene has in a new skeletal muscle network irrespective of the hub gene's original module members in PBMC module.

Figure 48. GS dynamics of the least hub genes in different skeletal muscle conditions. Absolute GS of the selected genes from PBMC were examined in four different conditions. GS: gene significance. v.s. PBMC (** $p < 0.01$, *** $p < 0.001$), v.s. acute resistance (24 hr, day 4) (# $p < 0.05$, ### $p < 0.001$), v.s. acute resistance (post 4 hr) (‡‡ $p < 0.01$). ANOVA with Tukey.

Figure 49. GS dynamics of the hub genes (the least preserved module, magenta) in different skeletal muscle conditions. v.s. PBMC (*** $p < 0.001$). ANOVA with Tukey.

Figure 50. GS dynamics of the hub genes (the most preserved module, turquoise) in different skeletal muscle conditions. v.s. PBMC (** $p < 0.01$, *** $p < 0.001$), v.s. acute resistance (24 hr, day 4) (# $p < 0.05$). ANOVA with Tukey.

Figure 51. kME dynamics of the least hub genes in different skeletal muscle conditions. kME of the selected genes from PBMC were examined in three different conditions. Three, but not four, skeletal muscle conditions are shown here because Resistance (post

24 hr, day 2) and Resistance (post 24 hr, day 4) share the same network topology and their kMEs are the same unlike GS values. kME: eigengene-based connectivity. v.s. PBMC (*** $p < 0.001$).

Figure 52. kME dynamics of the hub genes (the least preserved module, magenta) in different skeletal muscle conditions. v.s. PBMC (* $p < 0.05$, ** $p < 0.01$).

Figure 53. kME dynamics of the hub genes (the most preserved module, turquoise) in different skeletal muscle conditions. v.s. PBMC (*** $p < 0.001$), v.s. acute resistance 24 hr (# $p < 0.05$).

Figure 54. Top 1000 edges connected to the turquoise/magenta hub genes preserved in acute resistance (post 24 hr). Node colors correspond to the module of the gene in acute resistance (post 24 hr). Edge colors represent the original module of the connected hub gene (turquoise or magenta) in PBMC. Node size and edge width indicate summed connectivity of the node and edge weight, respectively.

Figure 55. Cluster dendrogram based on the calculated adjacency matrix of acute resistance (post 4 hr). To identify the network modules in which highly co-expressed genes are clustered together ($r > 0.85$), dynamic hybrid tree cutting algorithm was used. Different module colors indicate different network modules.

Figure 56. Correlation between eigengene and exercise traits. To identify interesting modules, Pearson correlations between eigengene and exercise traits (comparison between pre and post 4 hr of exercise) were examined. Correlation between eigengene and age (or sex) is also shown to clarify that the significant correlation between the selected module eigengene and exercise trait (e.g., green module) is not a cofounding of age or sex effect. ME: Module Eigengene.

Figure 57. Correlation between module membership and absolute GS for the exercise trait in acute resistance (post 4 hr). Five modules (darkgrey, green, grey60, yellowgreen, white) were selected.

Figure 58-60. Preservation of module network of acute resistance (post 4 hr) in different conditions.

Module preservation was measured by $Z_{summary}$ and medianRank for the selected modules in acute resistance (post 4 hr). $Z_{summary}$ with less than 2 (shown in red line) indicates no preservation (Figure 58). $Z_{density}$ and relative preservation (medianRank density) focusing on network density (Figure 59). $Z_{connectivity}$ and relative preservation (medianRank connectivity) focusing on module member connectivity to the hub gene (Figure 60).

Figure 61-63. Network in the green module preserved between acute resistance (post 4 hr) and acute resistance (post 24 hr).

(Figure 61) Acute resistance (post 4 hr) network associated with the preserved network between acute resistance (post 4 hr) and acute resistance (post 24 hr). The genes are all related to TGF signaling, Insulin signaling, and MAPK signaling (all in KEGG pathways), visualized by Cytoscape and only the nodes connected to at least one edge

(edge weight > 0.03) were used to avoid complexity. A darker and thicker edge indicates a larger edge weight. Green and red color nodes indicate negative and positive correlations in GS values of acute resistance (post 4 hr), respectively.

(Figure 62) Acute resistance (post 24 hr, day 2) associated with the preserved network as described in Figure.6A. The same parameters were applied for visualization of edges and nodes except that only the nodes connected to at least one edge (edge weight > 0.05) were used to avoid complexity.

(Figure 63) Acute resistance (post 24 hr, day 4) associated with the preserved network as described in Figure.6A. As in Figure.6B, the same parameters were applied for visualization of edges and nodes. Acute resistance exercise was carried out on day 1 and day 3. VL was sampled on day 2 (post 24 hr, day 2) and on day 4 (post 24 hr, day 4).

Figure 64. Newly formed connections in acute resistance (post 24 hr) to the strongly preserved 8 genes between acute resistance (post 4 hr) and (post 24 hr).

Due to the well-preserved connection and dynamic changes in GS across time, the small sub-network (*TNFRSF1A*, *MYC*, *IL1R1*, *GADD45A*, *GADD45B*, *GADD45G*, *CRK*, and

NR4A1) was focused and the network genes connected to these 8 genes in acute resistance (post 24 hr) was visualized by Cytoscape. Top 500 strongest edges were shown. A darker and thicker edge indicates a larger edge weight. Node size and color indicate summed connection strength of the node to other nodes and the corresponding module in post 24 hr network, respectively.

Figure 65. GS values of top 500 edge nodes were compared between post 24 hr day 2 and day 4. The top 500 strongest edges were selected from the network in acute resistance (post 24 hr) connected to the genes (*TNFRSF1A, MYC, IL1R1, GADD45A, GADD45B, GADD45G, CRK,* and *NR4A1*). Kolmogorov-Smirnov test was used for statistics.

Figure 66,67. Network in the green module preserved between acute resistance (post 4 hr) and acute endurance (post 2.5 hr and 5 hr).

(Figure 66) Acute resistance (post 4 hr) network associated with the preserved network between acute resistance (post 4 hr) and acute endurance (post 2.5 hr and 5 hr). The

genes are all related to TGF signaling, Insulin signaling, and MAPK signaling (all in KEGG pathways), visualized by Cytoscape. A darker and thicker edge indicates a larger edge weight. Green and red color nodes indicate negative and positive correlations in GS values of acute resistance (post 4 hr), respectively.

(Figure 67) Acute endurance (post 2.5 hr and 5 hr) associated with the preserved network.

FIGURES AND TABLES

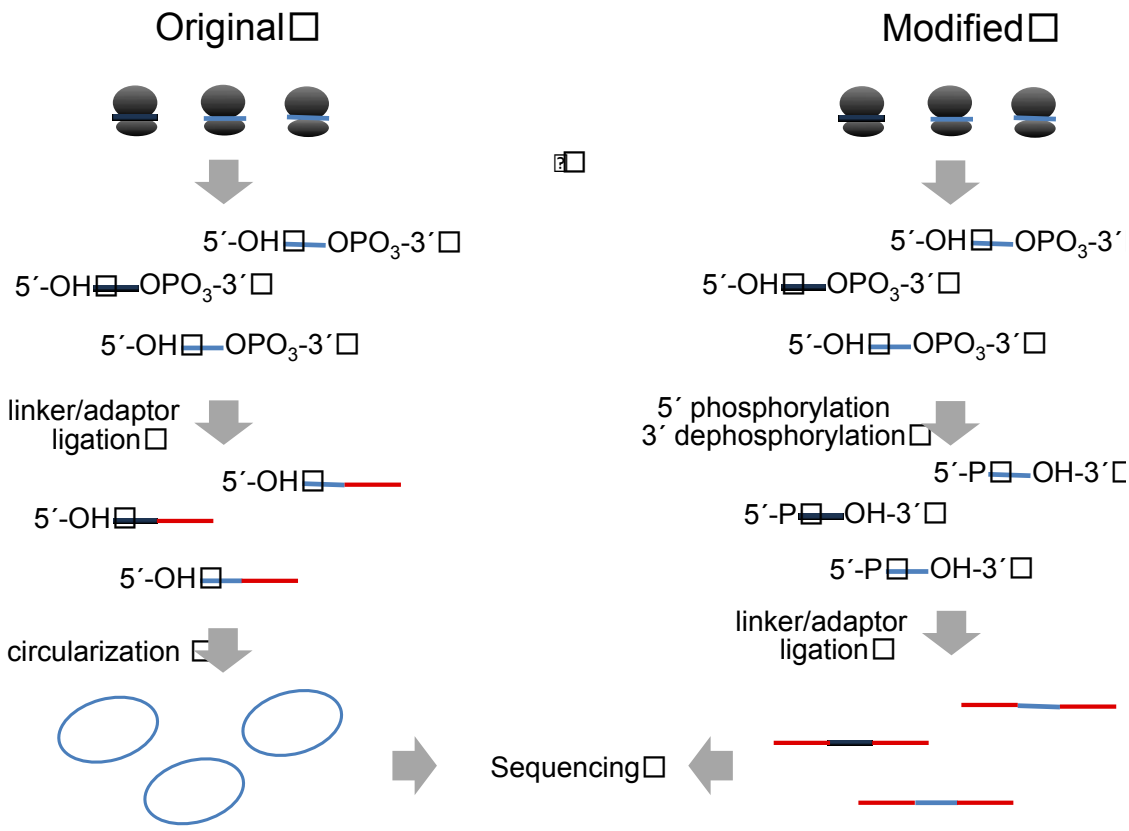

Original
Modified
5′-OH
OPO3-3′
linker/adaptor ligation
5′ phosphorylation
3′ dephosphorylation
5′-P
OH-3′
circularization
Sequencing

Figure 1.

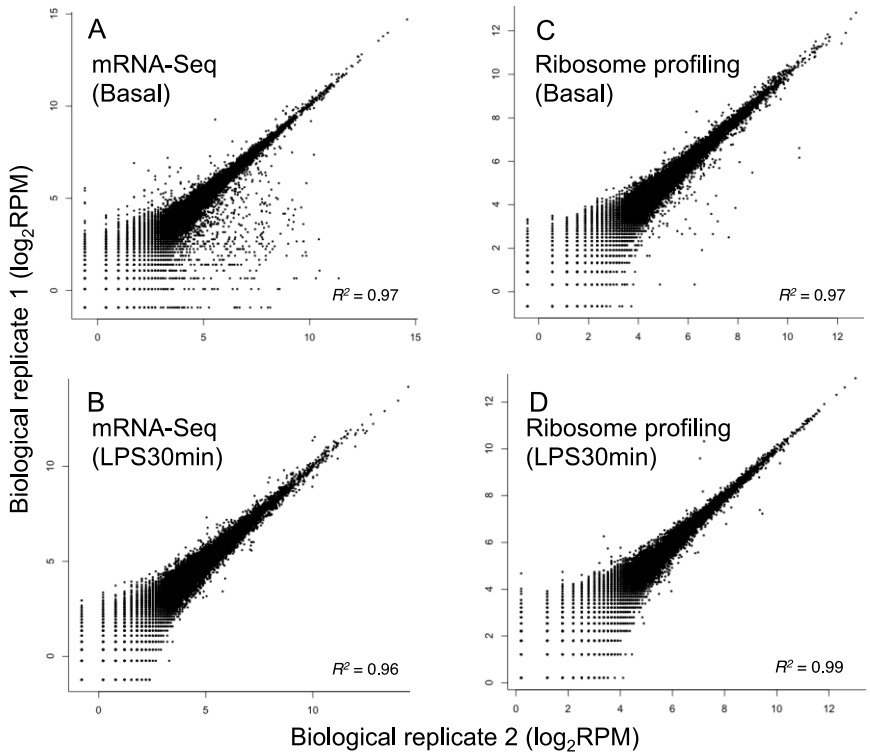

A
mRNA-Seq
(Basal)
R² = 0.97
C
Ribosome profiling
(Basal)
R² = 0.97
B
mRNA-Seq
(LPS30min)
R² = 0.96
D
Ribosome profiling
(LPS30min)
R² = 0.99
Biological replicate 1 (log2RPM)
Biological replicate 2 (log2RPM)

Figure 2.

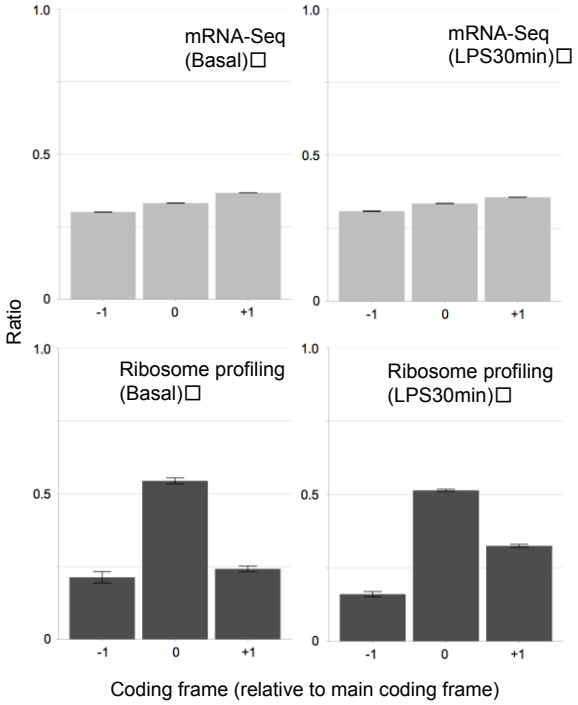


Figure 3.

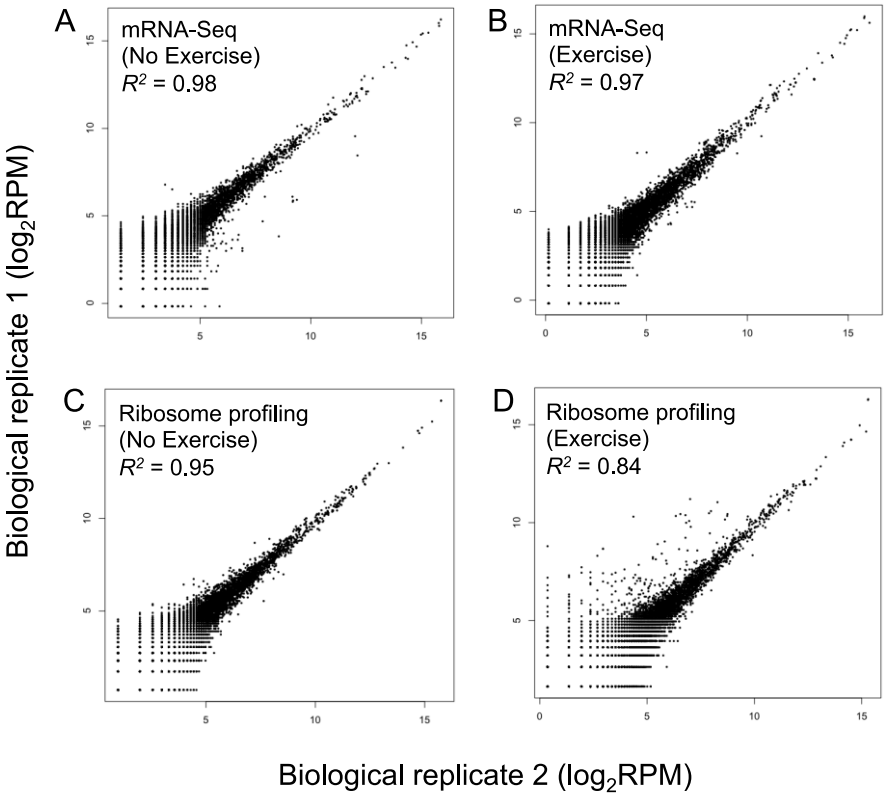


Figure 4.

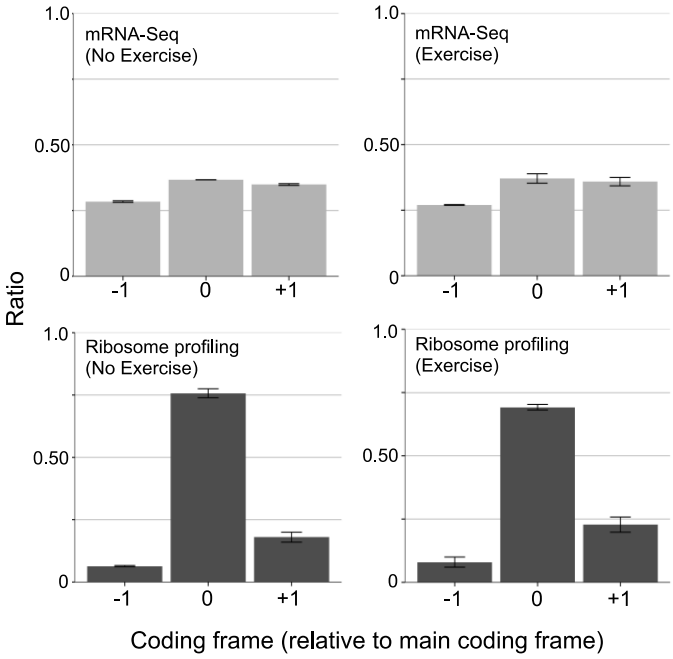


Figure 5.

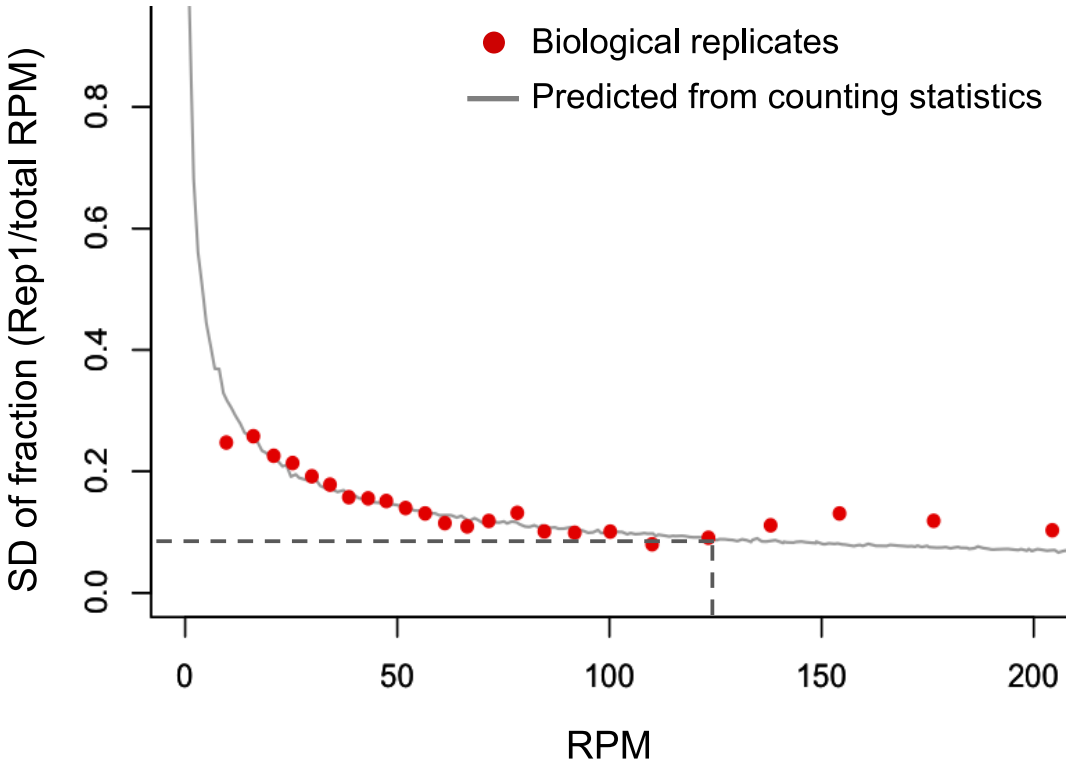


Figure 6.

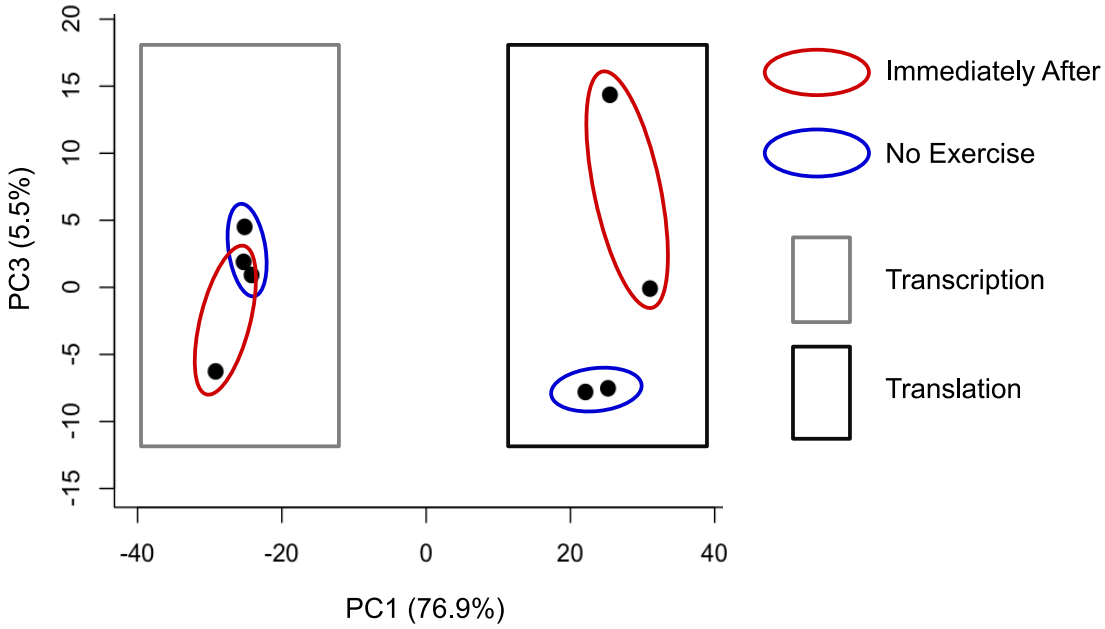


Figure 7.

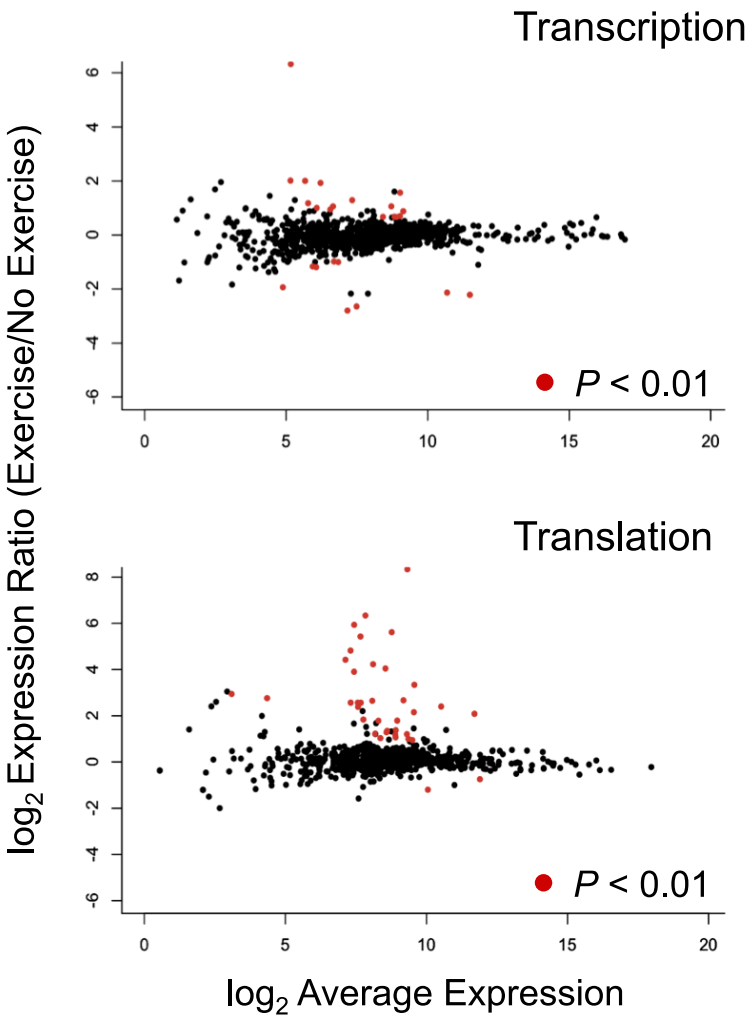

Transcription
Translation
P < 0.01
P < 0.01
log2 Expression Ratio (Exercise/No Exercise)
log2 Average Expression

Figure 8.

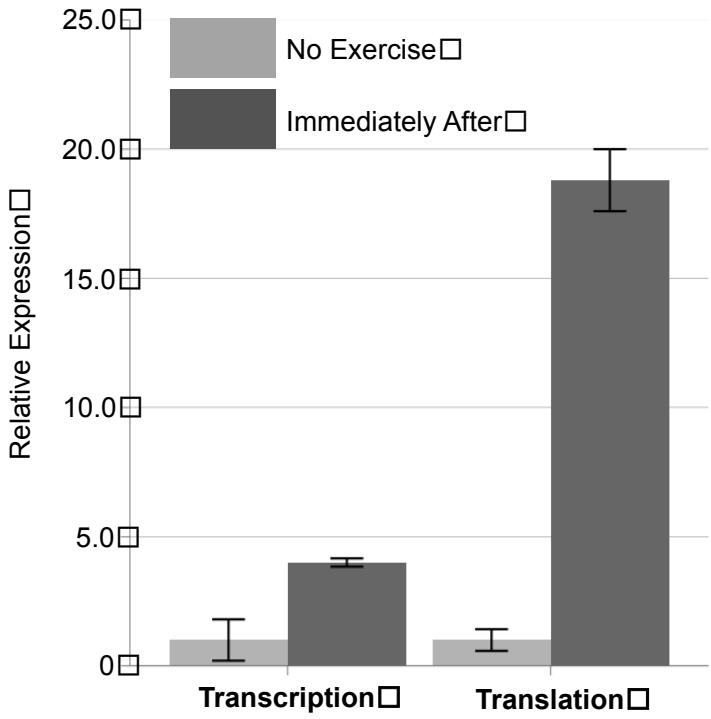


Figure 9.

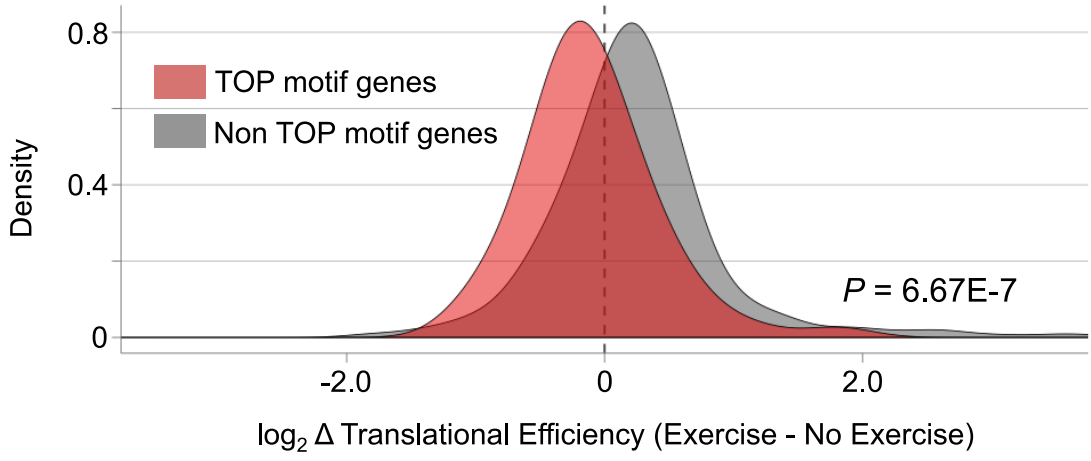


Figure 10.

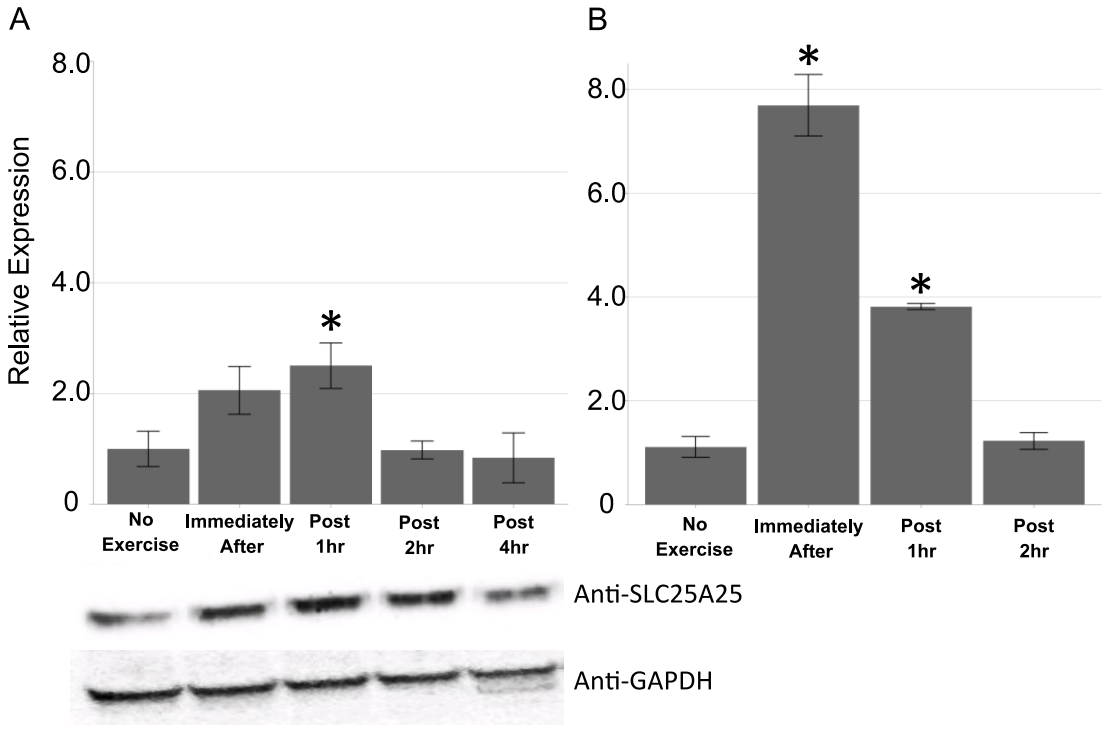


Figure 11.

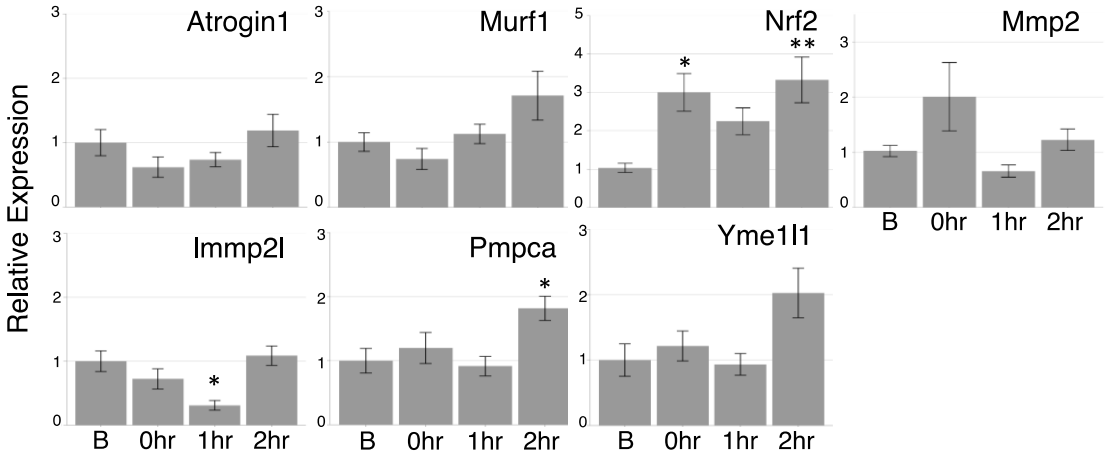


Figure 12.

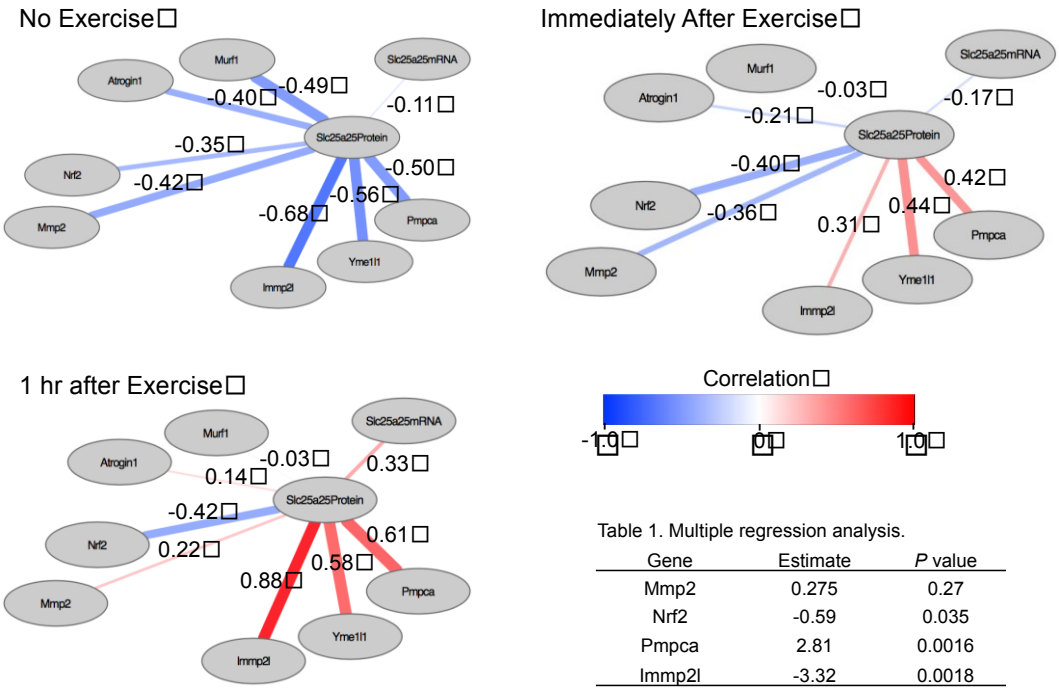


Table 1. Multiple regression analysis.

Gene	Estimate	*P* value
Mmp2	0.275	0.27
Nrf2	-0.59	0.035
Pmpca	2.81	0.0016
Immp2l	-3.32	0.0018

Figure 13.

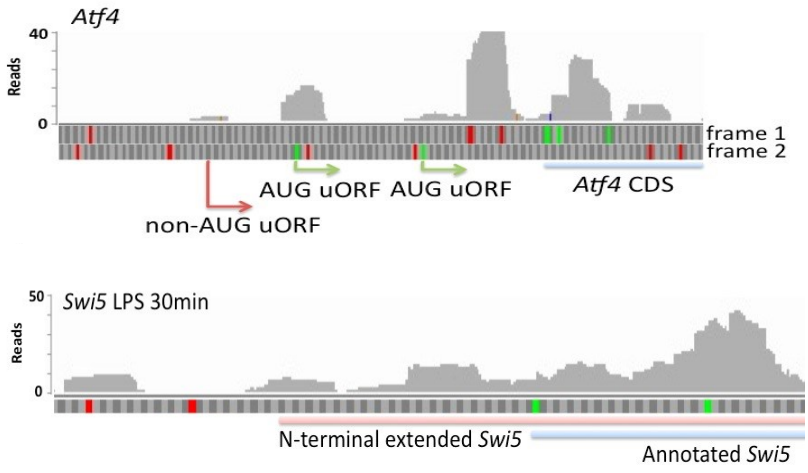


Figure 14.

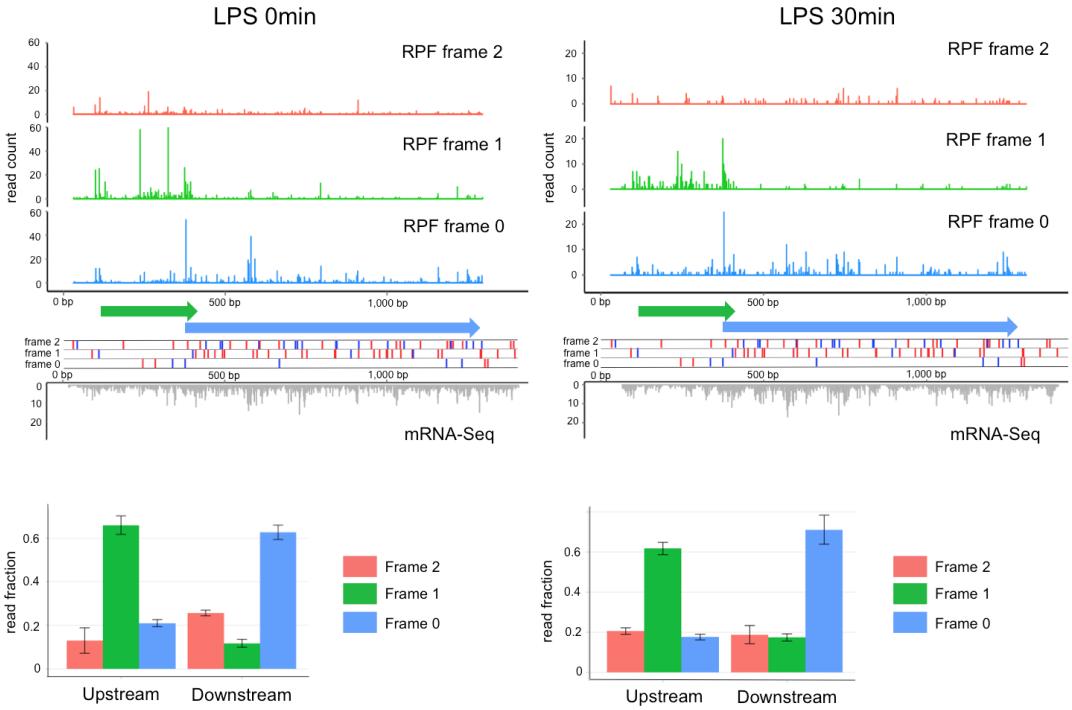


Figure 15.

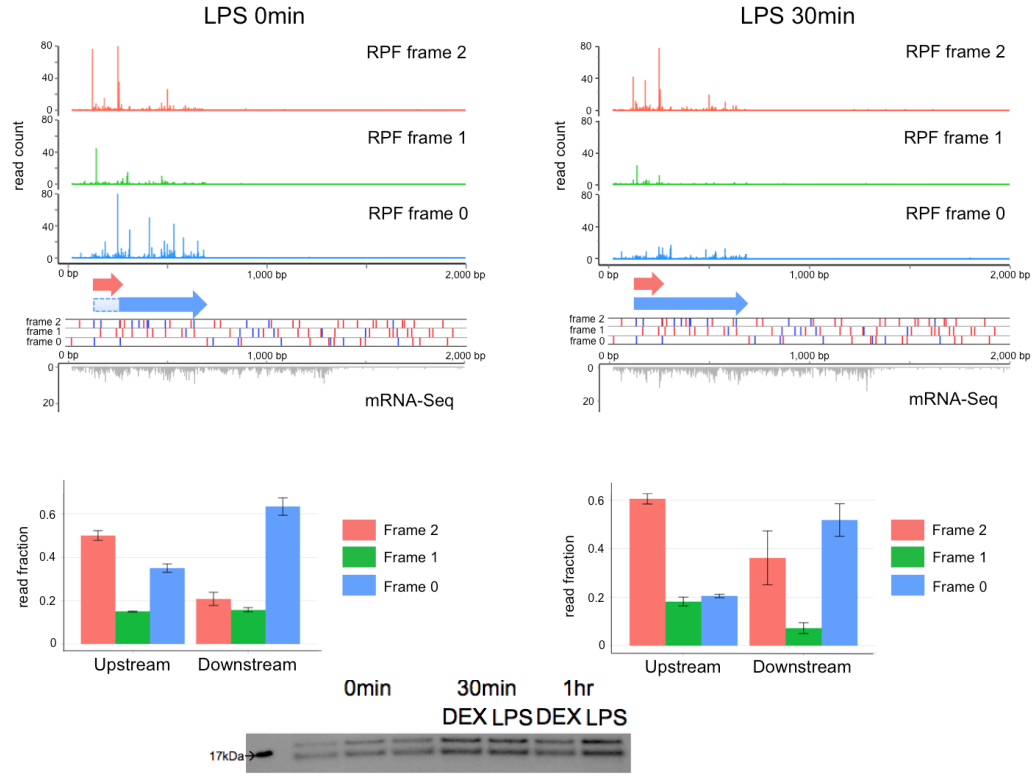


Figure 16.

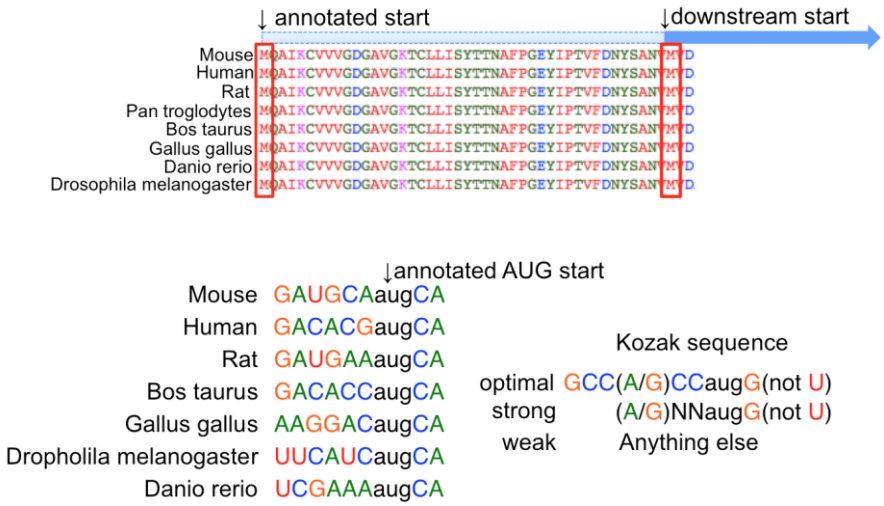


Figure 17.

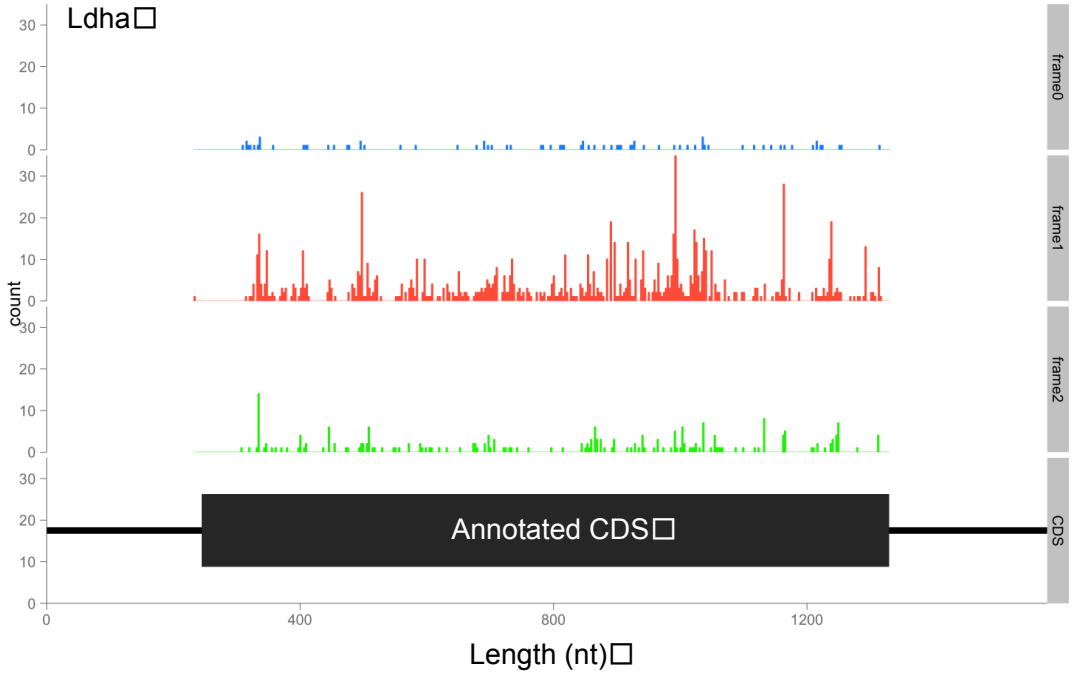


Figure 18.

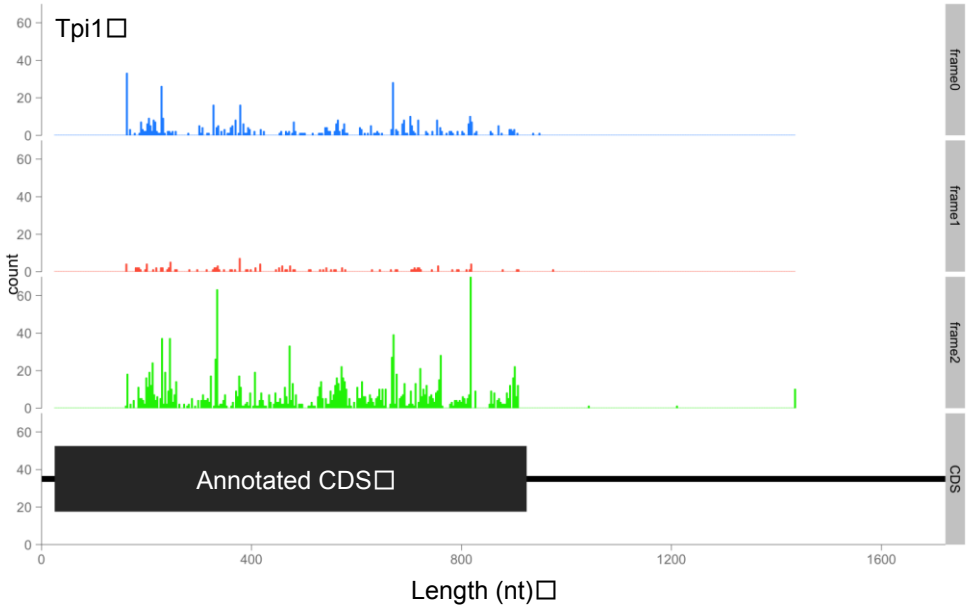


Figure 19.

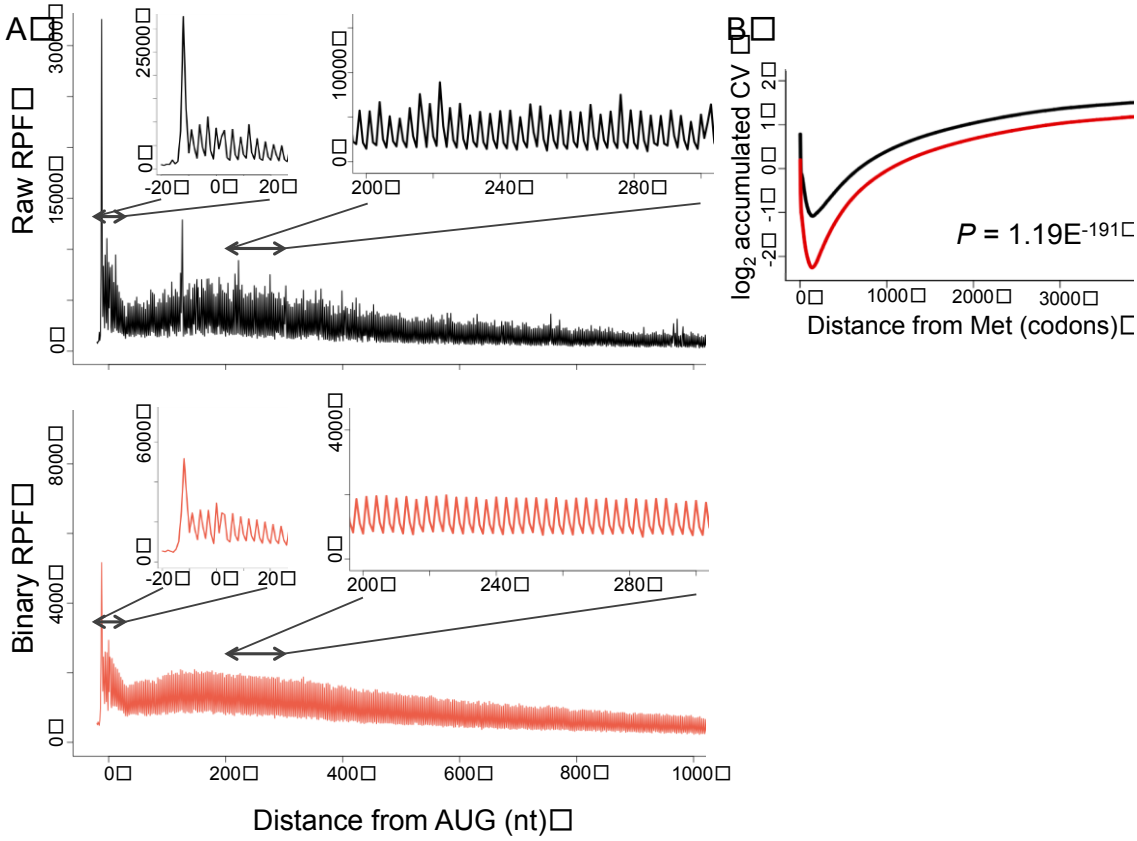


Figure 20.

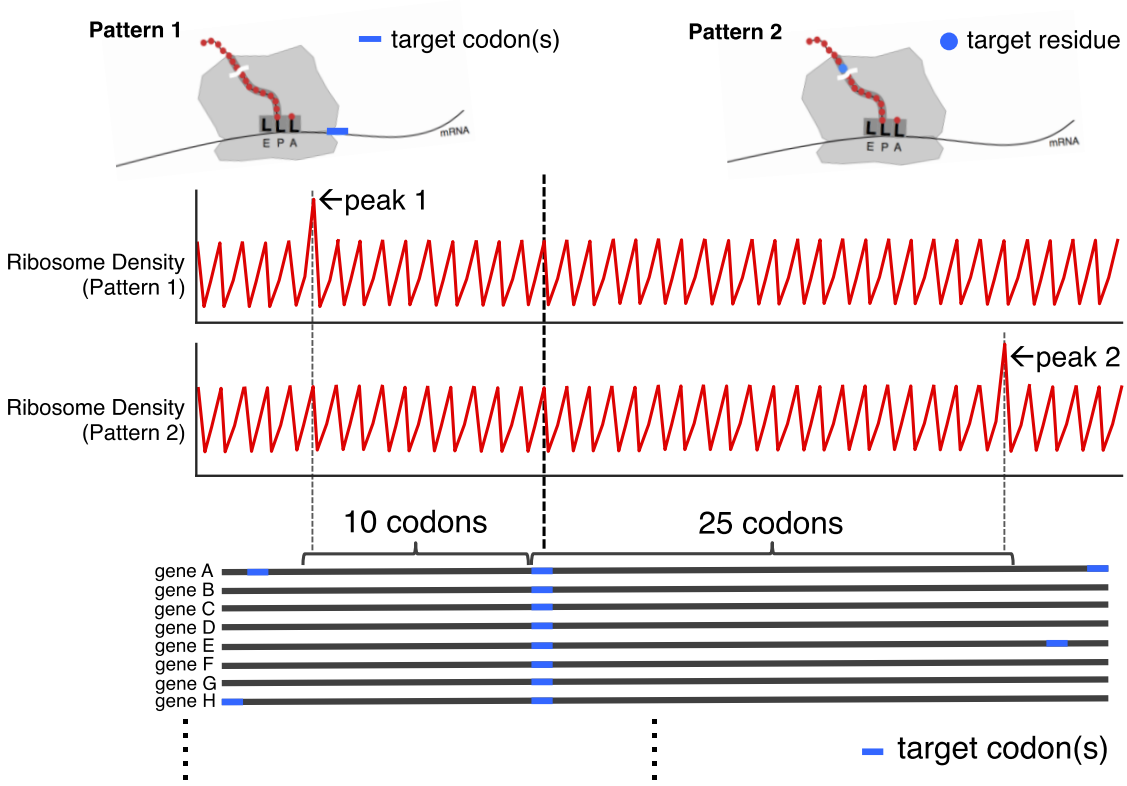

Pattern 1
target codon(s)
Pattern 2
target residue
E P A
mRNA
peak 1
Ribosome Density
(Pattern 1)
peak 2
Ribosome Density
(Pattern 2)
10 codons
25 codons
gene A
gene B
gene C
gene D
gene E
gene F
gene G
gene H
target codon(s)

Figure 21.

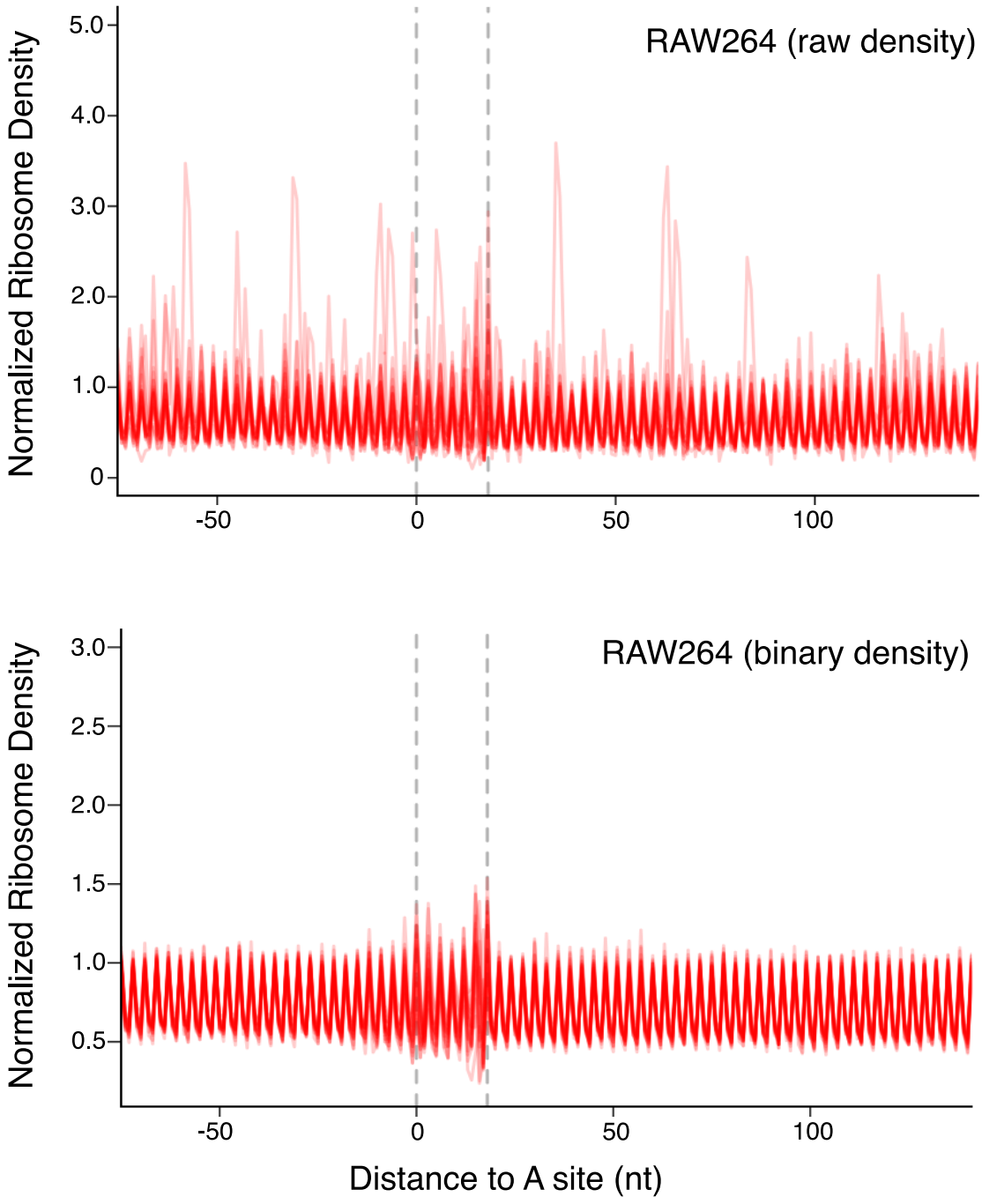

RAW264 (raw density)
Normalized Ribosome Density
5.0
4.0
3.0
2.0
1.0
0
-50
0
50
100
RAW264 (binary density)
Normalized Ribosome Density
3.0
2.5
2.0
1.5
1.0
0.5
-50
0
50
100
Distance to A site (nt)

Figure 22.

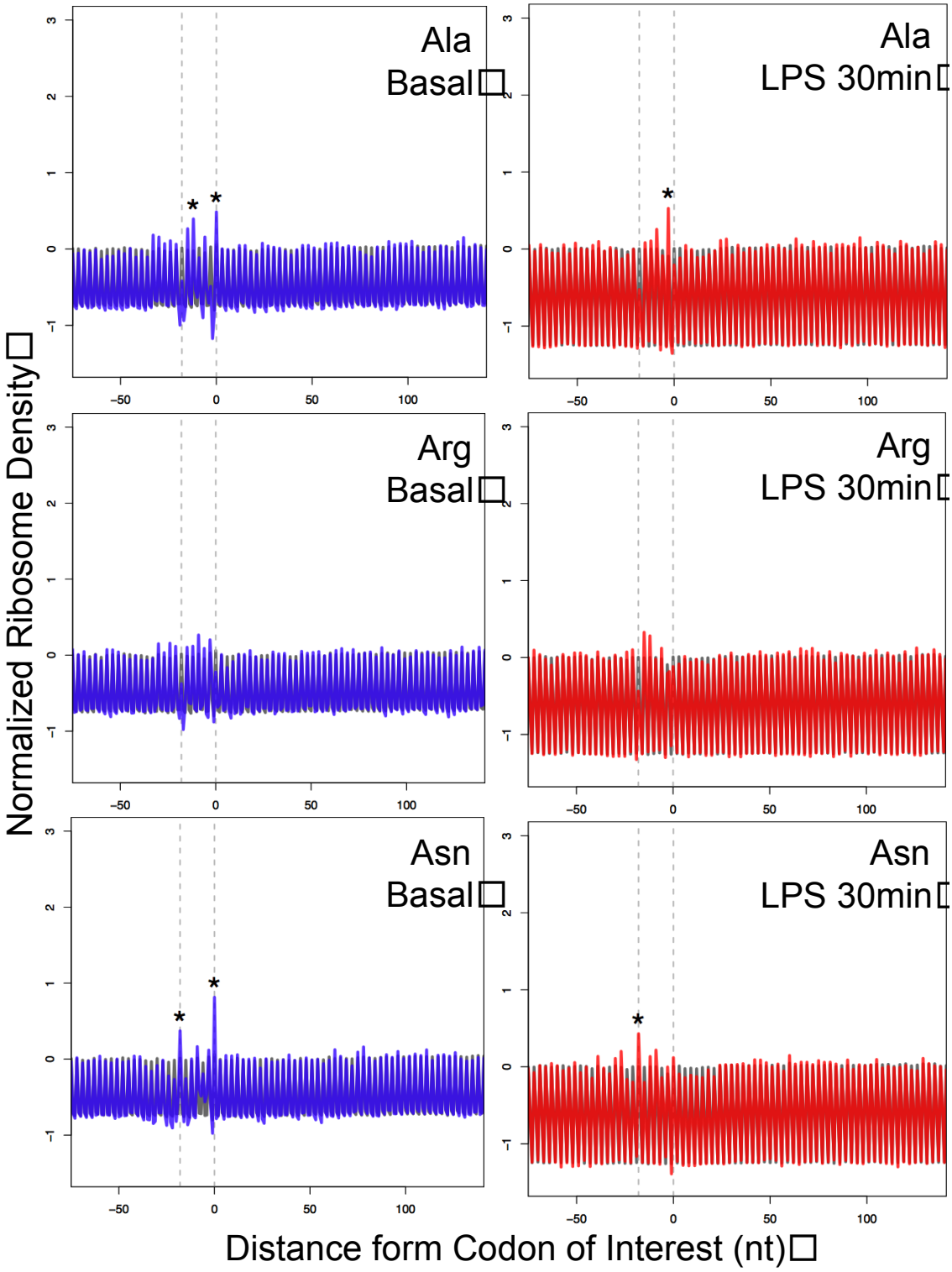

Ala
Basal
Ala
LPS 30min
Arg
Basal
Arg
LPS 30min
Asn
Basal
Asn
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 23.

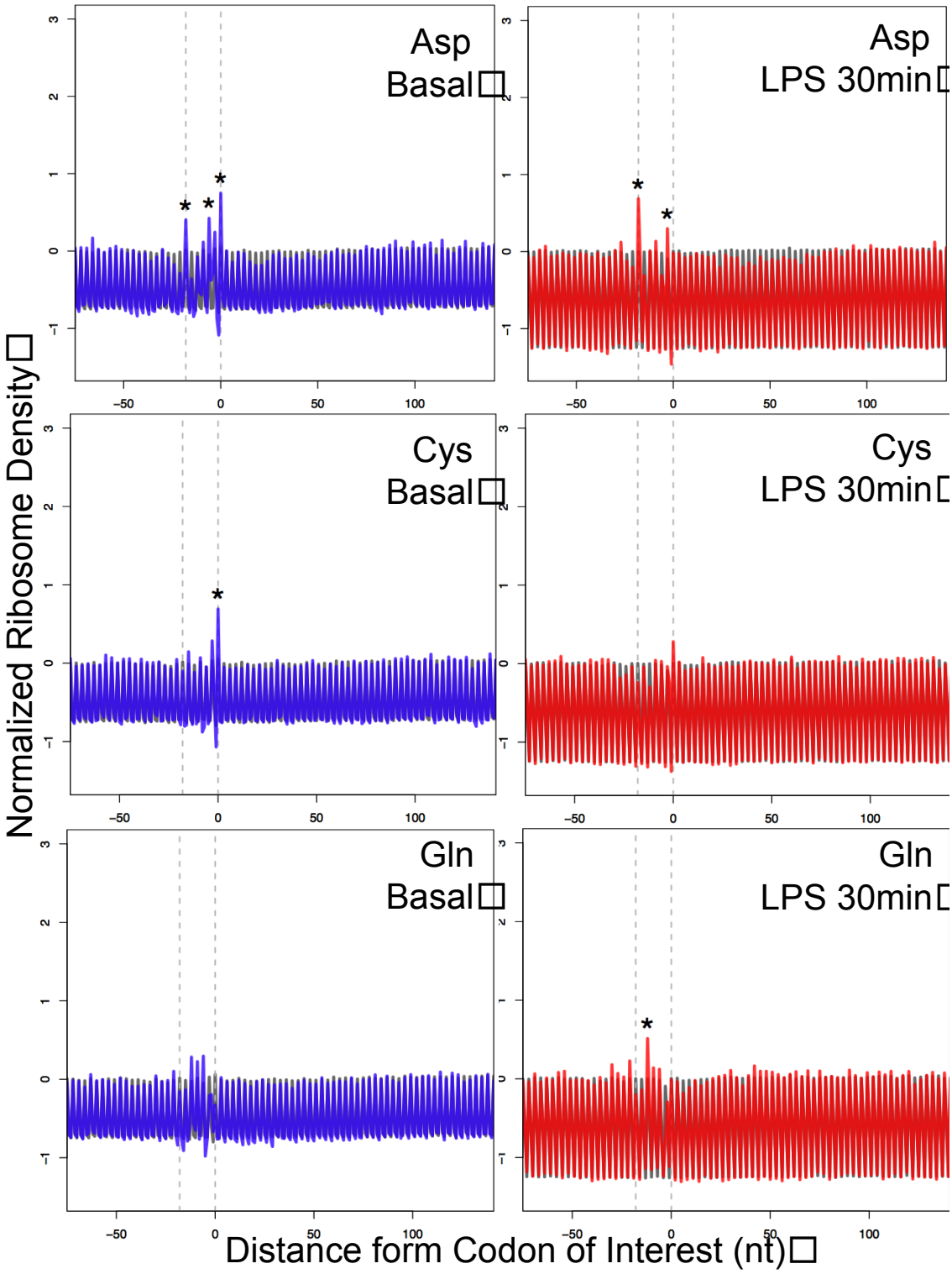

Asp
Basal
Asp
LPS 30min
Cys
Basal
Cys
LPS 30min
Gln
Basal
Gln
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 24.

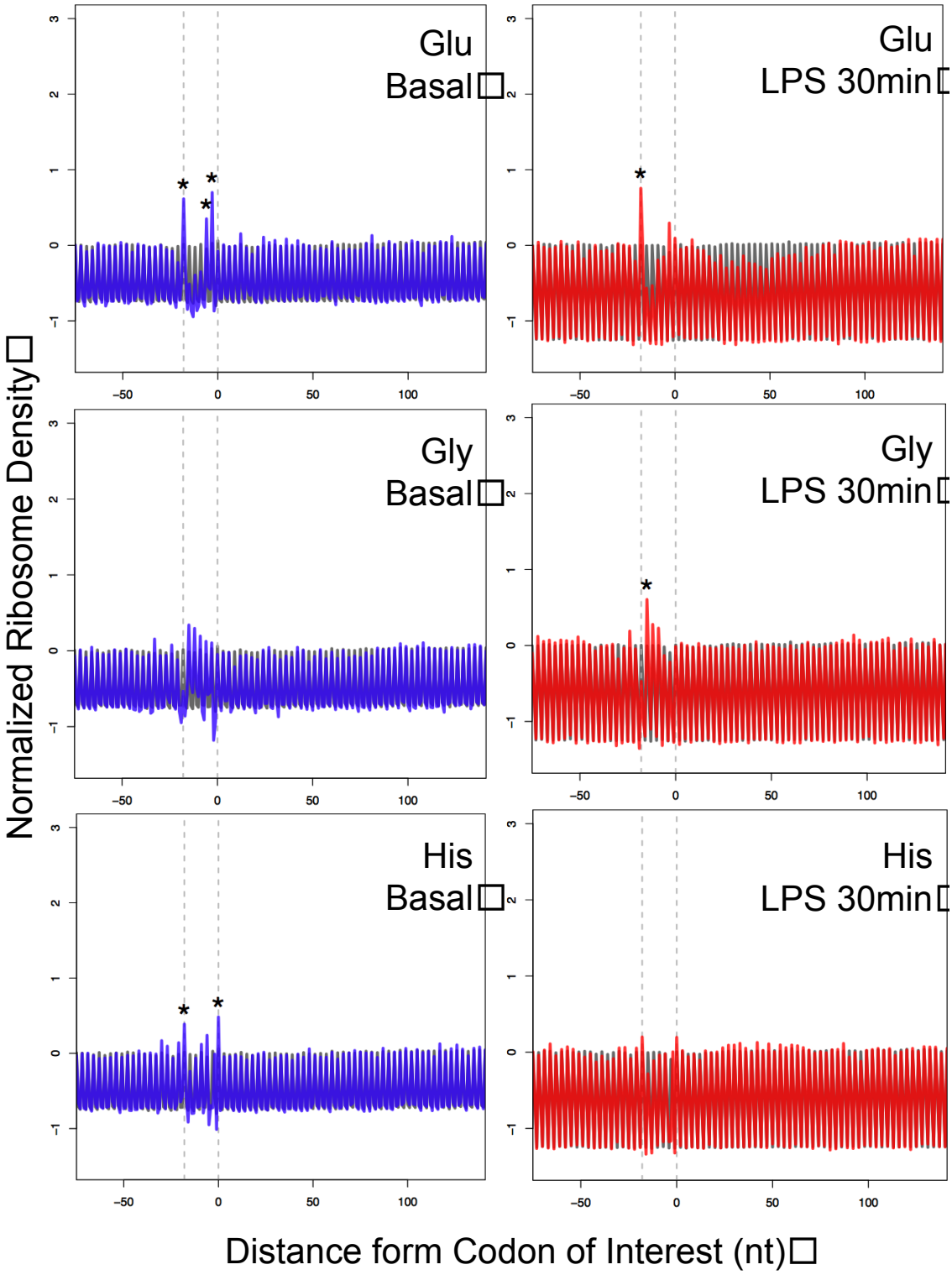

Glu
Basal
Glu
LPS 30min
Gly
Basal
Gly
LPS 30min
His
Basal
His
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 25.

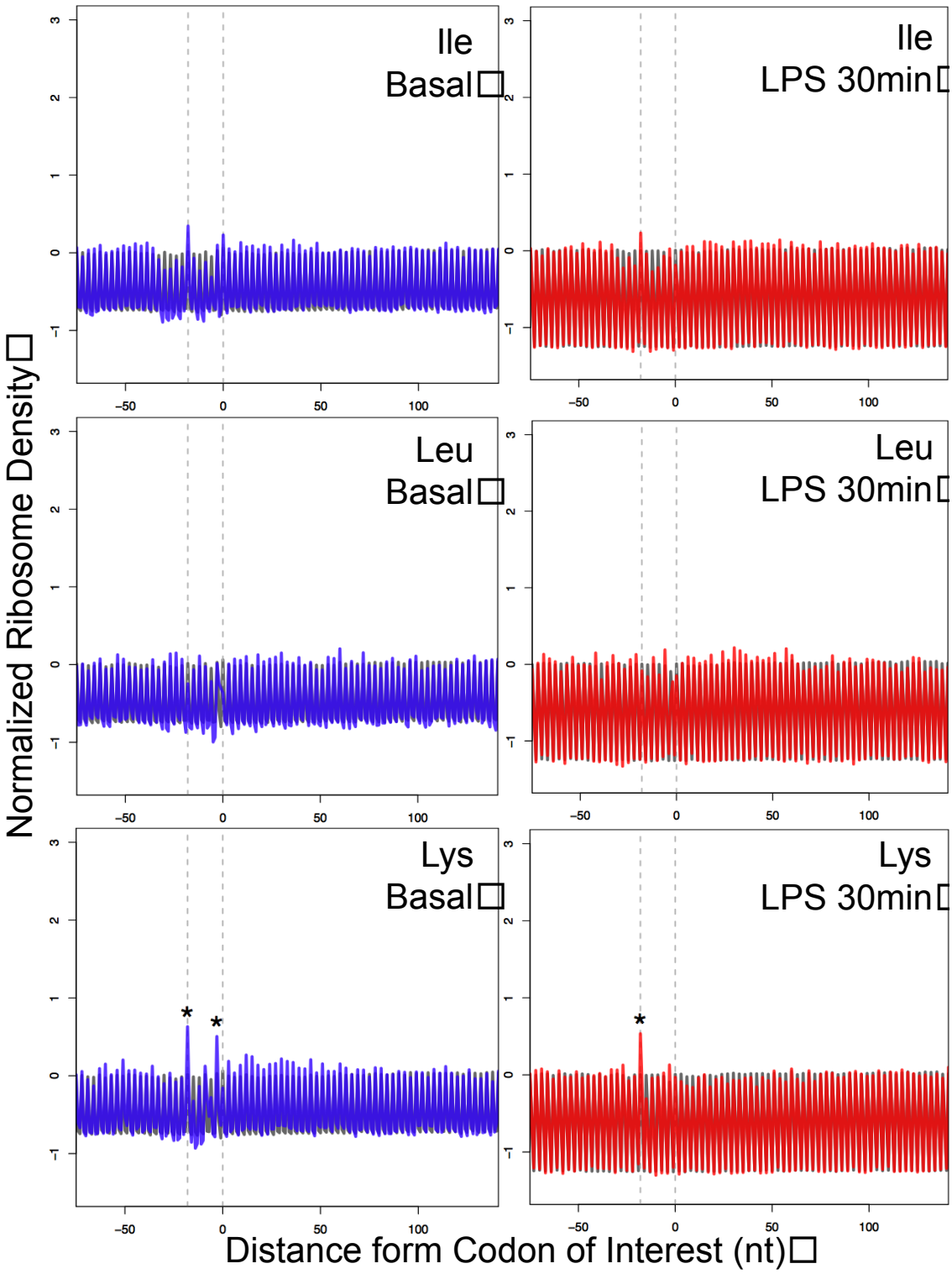

Ile
Basal
Ile
LPS 30min
Leu
Basal
Leu
LPS 30min
Lys
Basal
Lys
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 26.

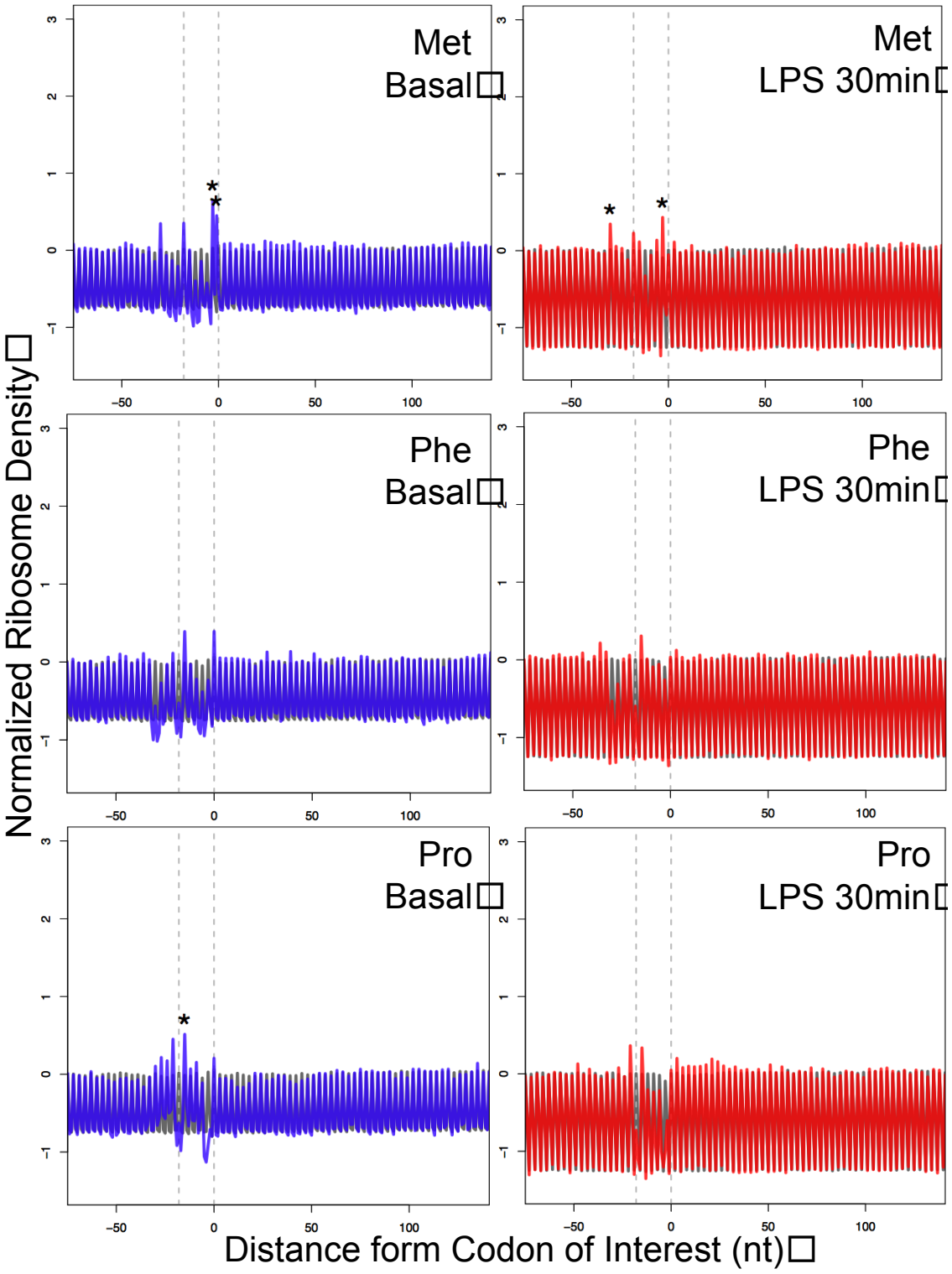

Met
Basal
Met
LPS 30min
Phe
Basal
Phe
LPS 30min
Pro
Basal
Pro
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 27.

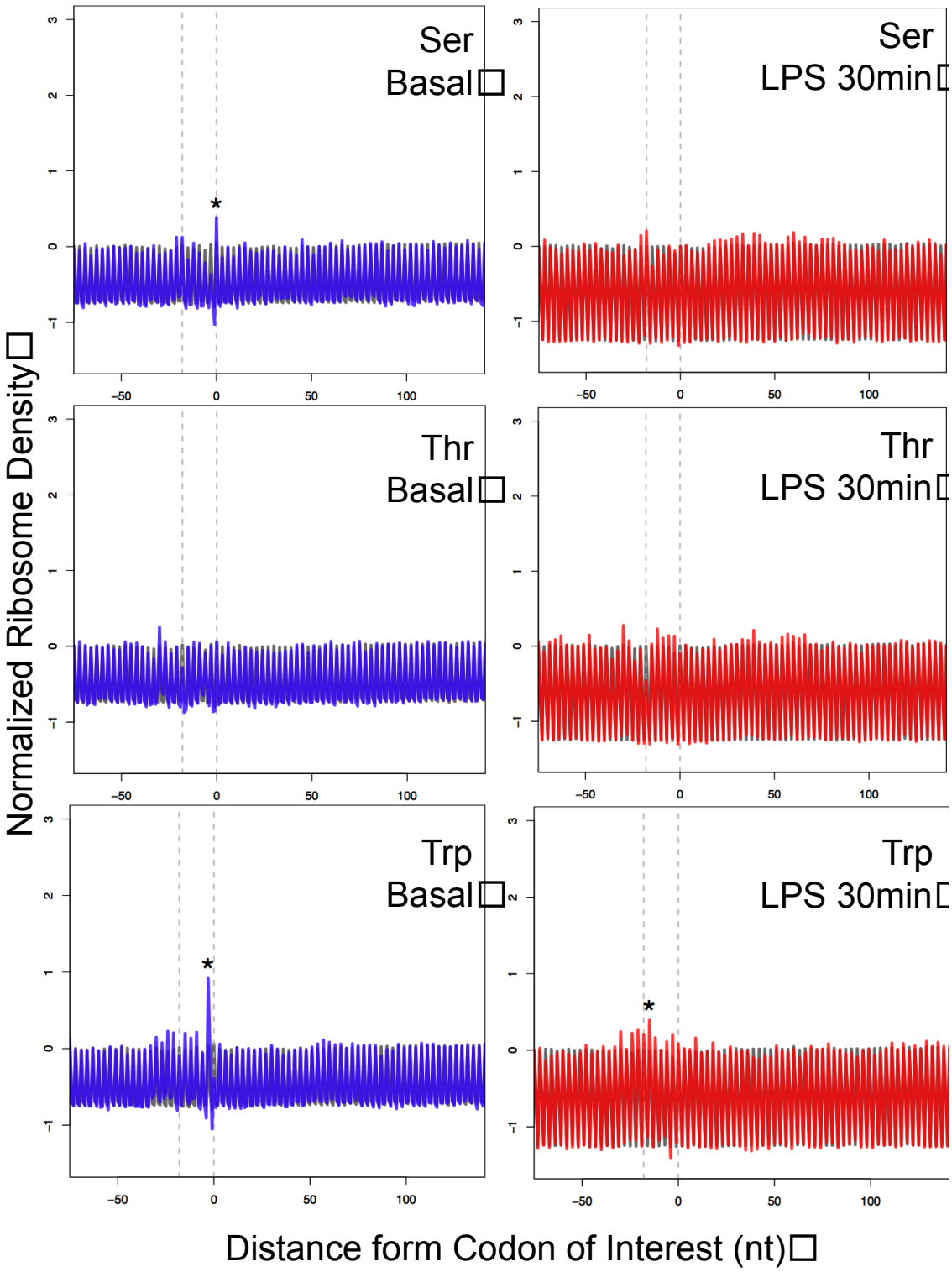

Ser
Basal
Ser
LPS 30min
Thr
Basal
Thr
LPS 30min
Trp
Basal
Trp
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 28.

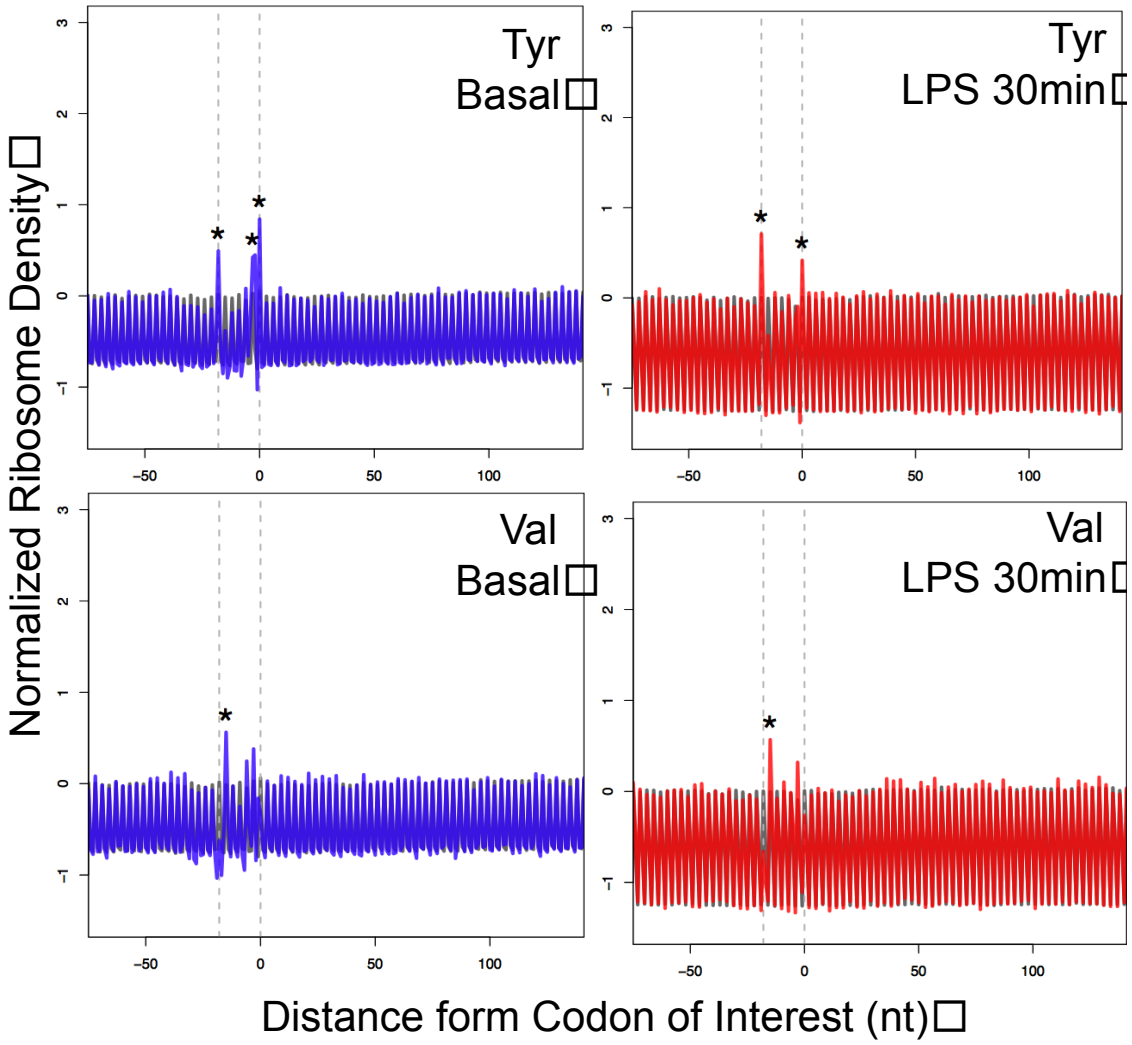

Tyr
Basal
Tyr
LPS 30min
Val
Basal
Val
LPS 30min
Normalized Ribosome Density
Distance form Codon of Interest (nt)

Figure 29.

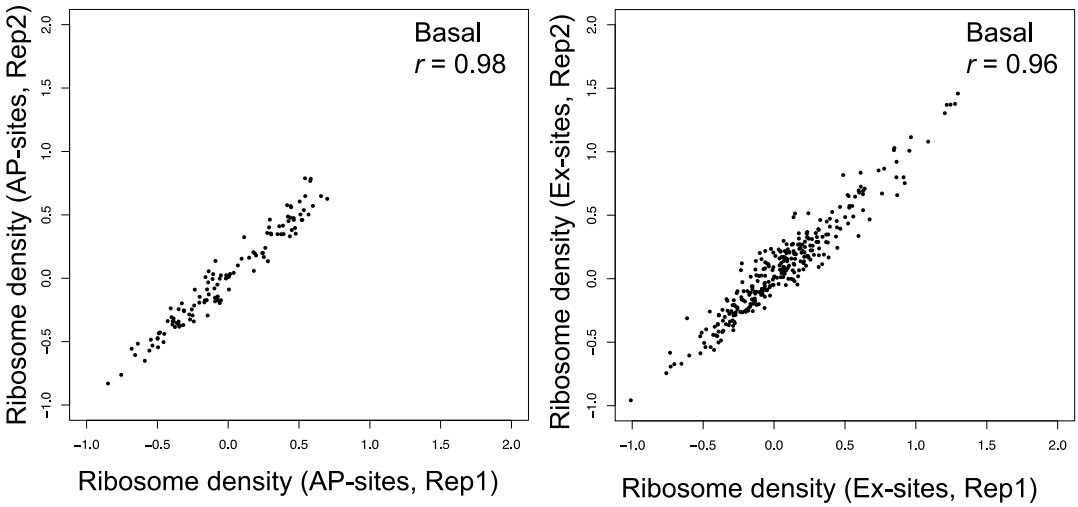


Figure 30.

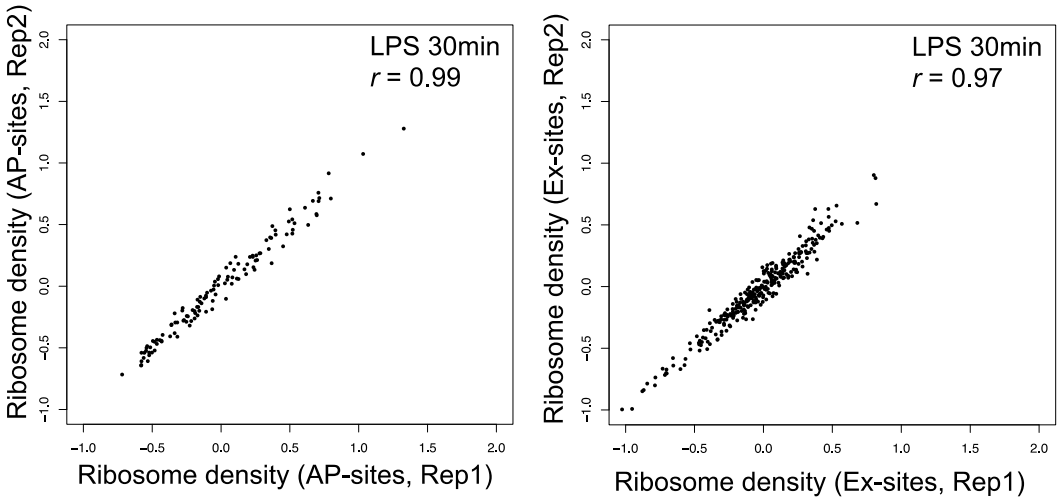


Figure 31

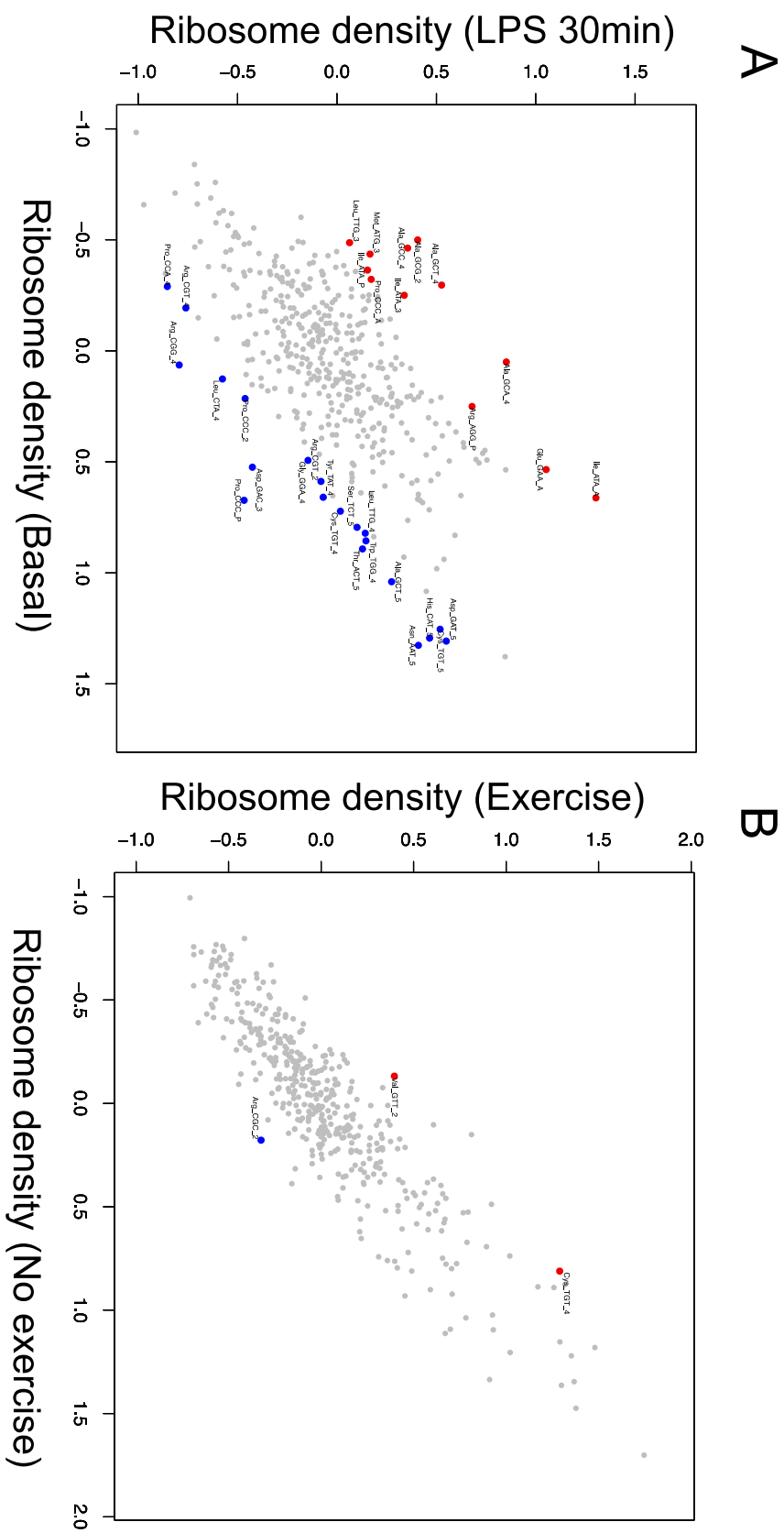

A
Ribosome density (LPS 30min)
Ribosome density (Basal)
B
Ribosome density (Exercise)
Ribosome density (No exercise)

Figure 32.

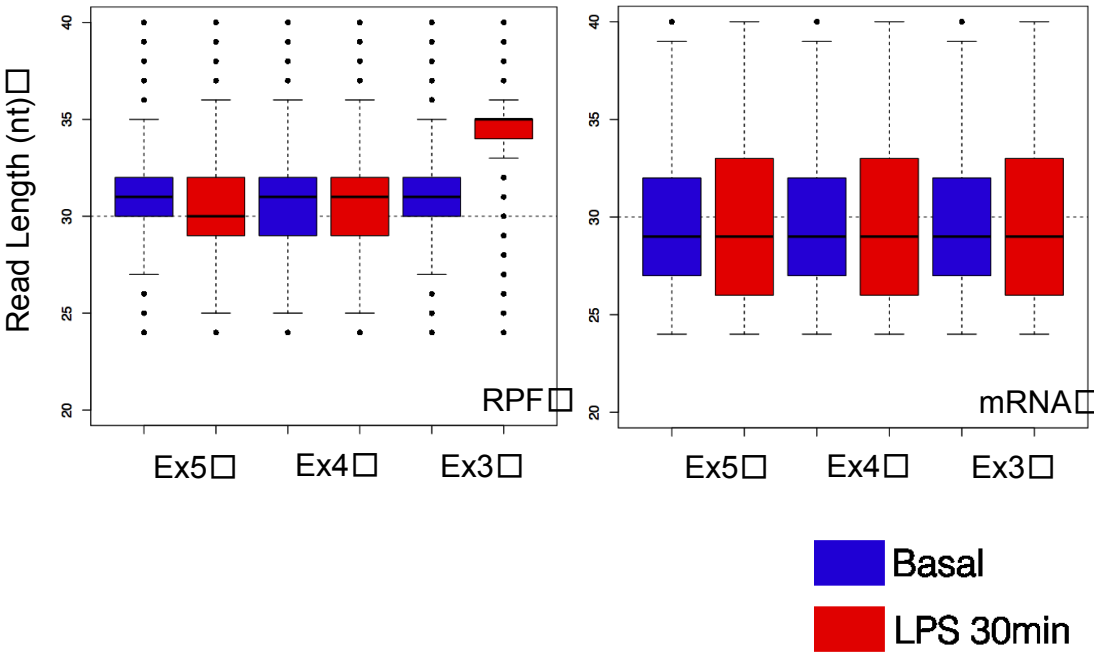

Read Length (nt)
20
25
30
35
40
RPF
mRNA
Ex5
Ex4
Ex3
Basal
LPS 30min

Figure 33.

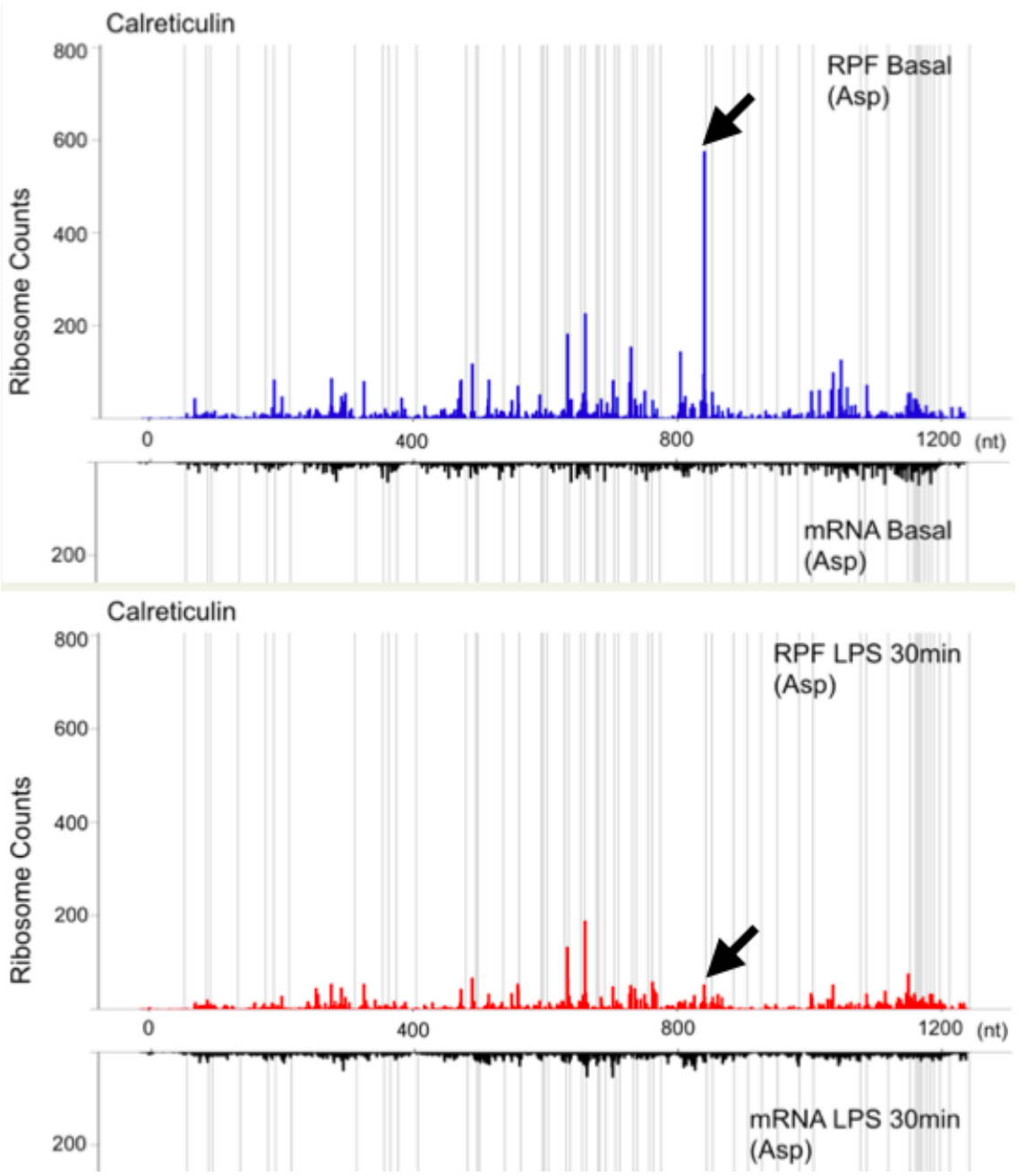

Calreticulin
Ribosome Counts
800
600
400
200
RPF Basal
(Asp)
0
400
800
1200
(nt)
200
mRNA Basal
(Asp)
Calreticulin
Ribosome Counts
800
600
400
200
RPF LPS 30min
(Asp)
0
400
800
1200
(nt)
200
mRNA LPS 30min
(Asp)

Figure 34.

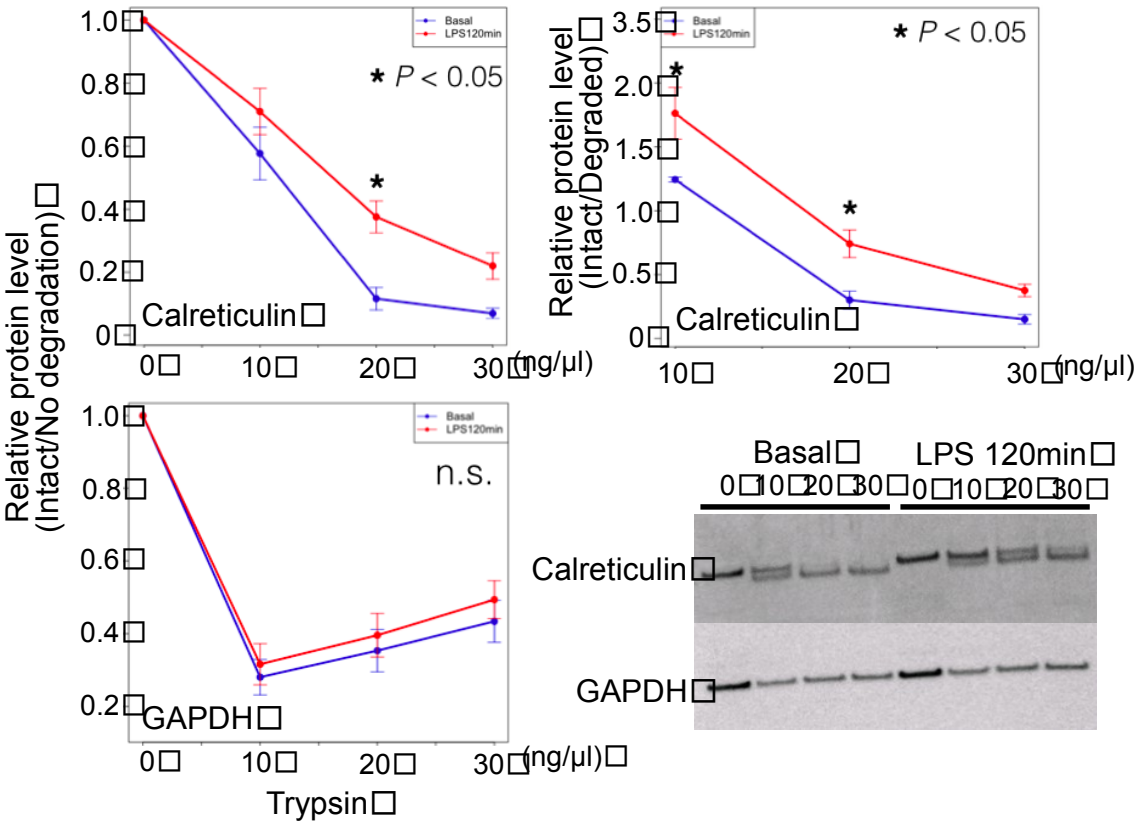

Relative protein level
(Intact/No degradation)
Basal
LPS120min
* P < 0.05
Calreticulin
(ng/μl)
Relative protein level
(Intact/Degraded)
n.s.
GAPDH
Trypsin
Basal
LPS 120min
0 10 20 30 0 10 20 30

Figure 35.

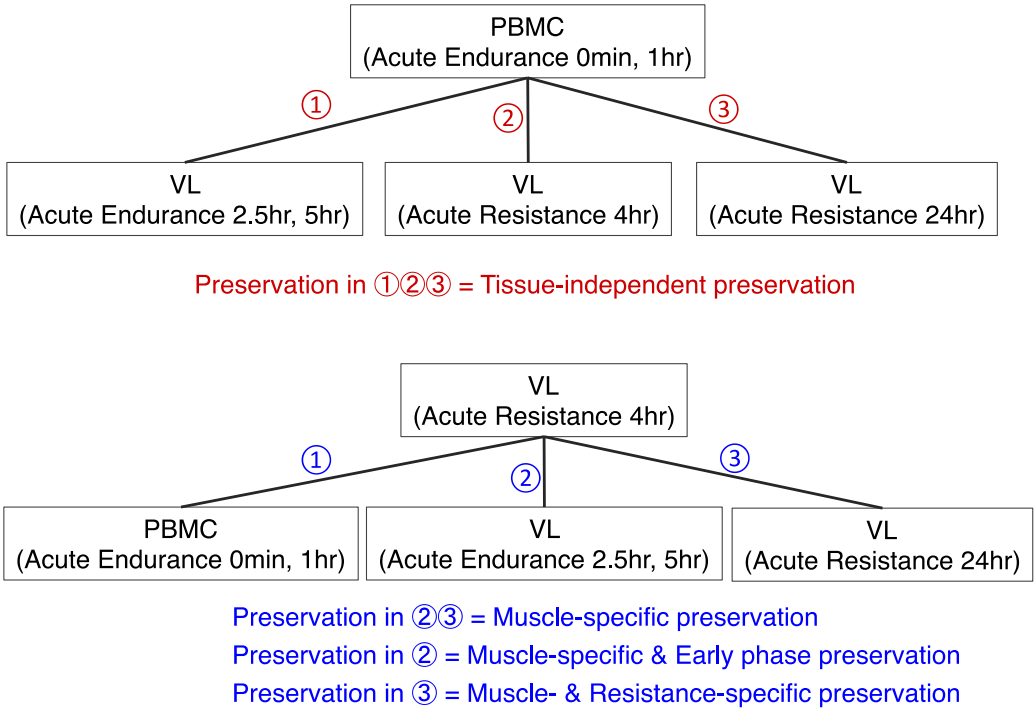


Figure 36.

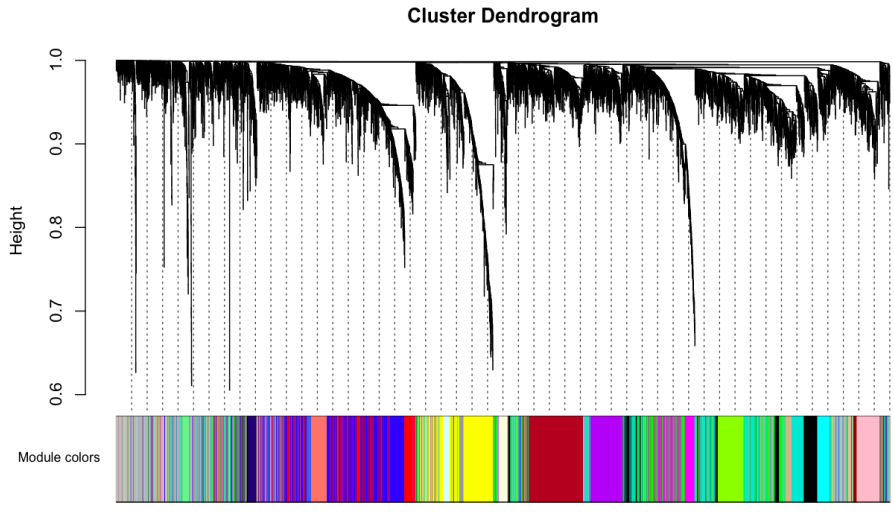


Figure 37.

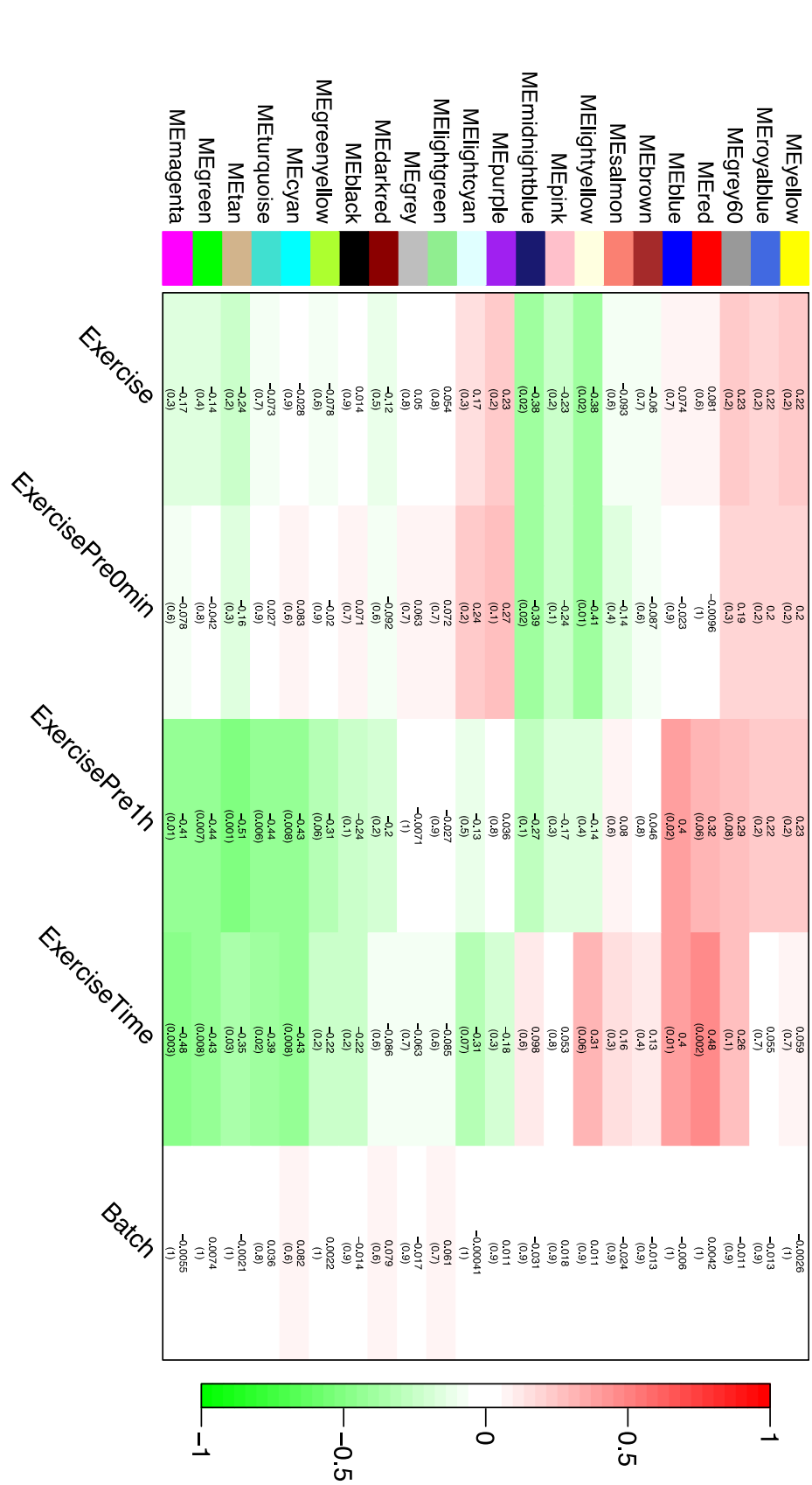

Module–trait relationships
MEyellow
MEroyalblue
MEgrey60
MEred
MEblue
MEbrown
MEsalmon
MElightyellow
MEpink
MEmidnightblue
MEpurple
MElightcyan
MElightgreen
MEgrey
MEdarkred
MEblack
MEgreenyellow
MEcyan
MEturquoise
MEtan
MEgreen
MEmagenta
Exercise
ExercisePre0min
ExercisePre1h
ExerciseTime
Batch
1
0.5
0
−0.5
−1

Figure 38.

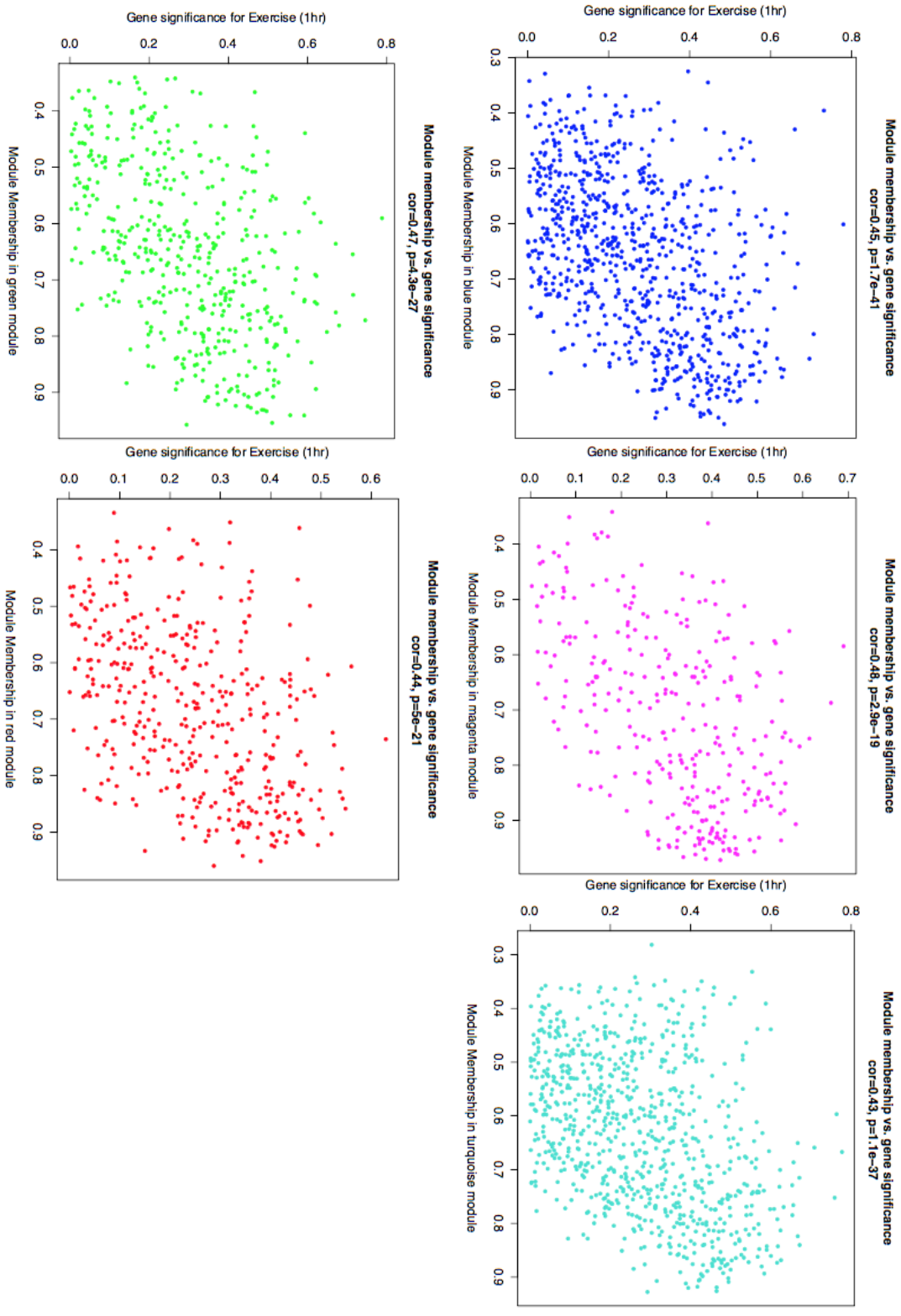

Gene significance for Exercise (1hr)
Module Membership in green module
Module membership vs. gene significance
cor=0.47, p=4.3e-27
Gene significance for Exercise (1hr)
Module Membership in blue module
Module membership vs. gene significance
cor=0.45, p=1.7e-41
Gene significance for Exercise (1hr)
Module Membership in red module
Module membership vs. gene significance
cor=0.44, p=5e-21
Gene significance for Exercise (1hr)
Module Membership in magenta module
Module membership vs. gene significance
cor=0.48, p=2.9e-19
Gene significance for Exercise (1hr)
Module Membership in turquoise module
Module membership vs. gene significance
cor=0.43, p=1.1e-37

Figure 39.

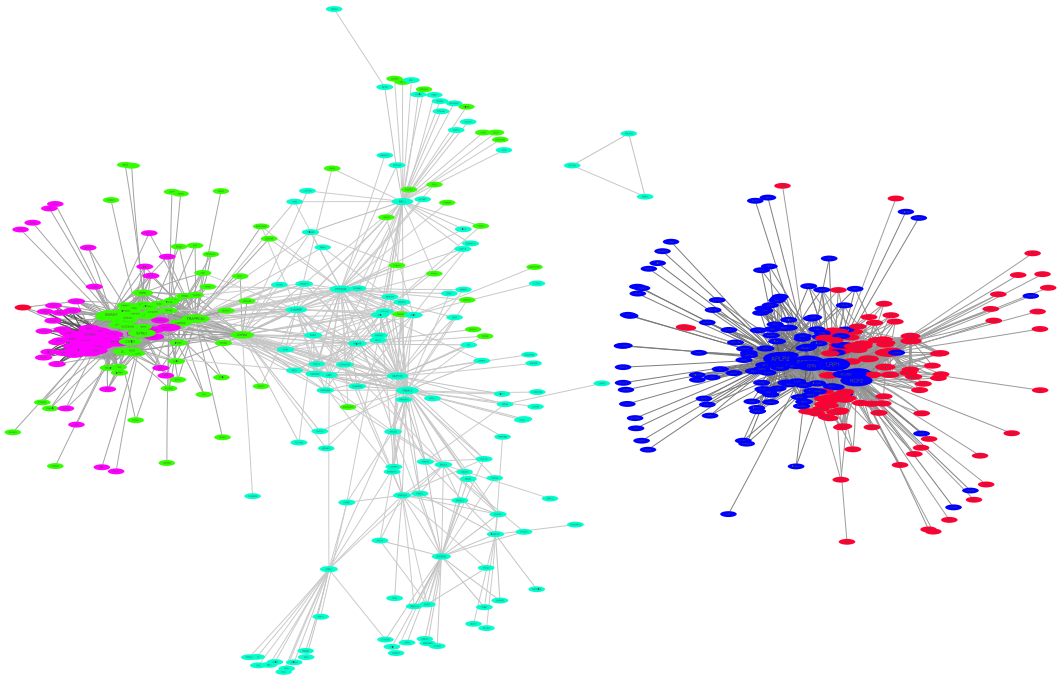


Figure 40.

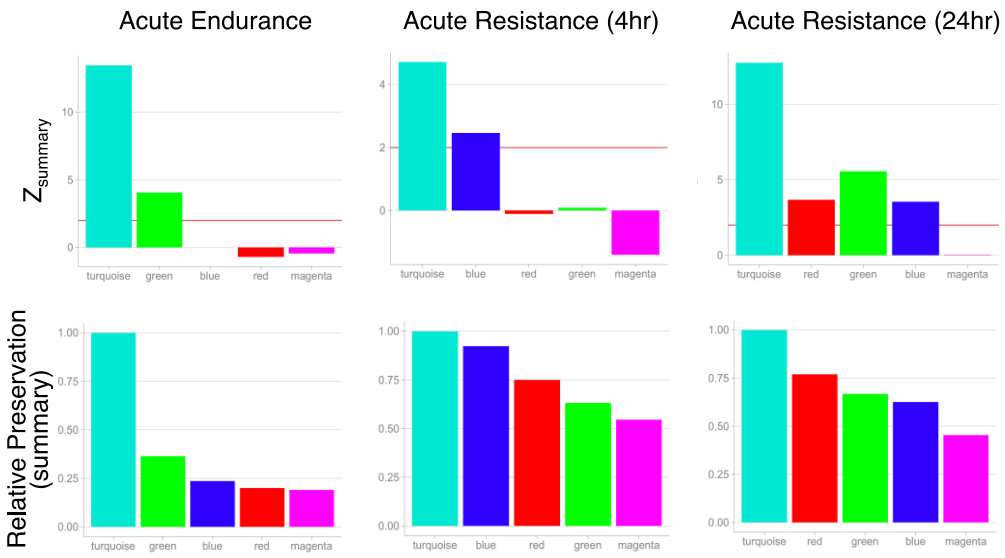


Figure 41.

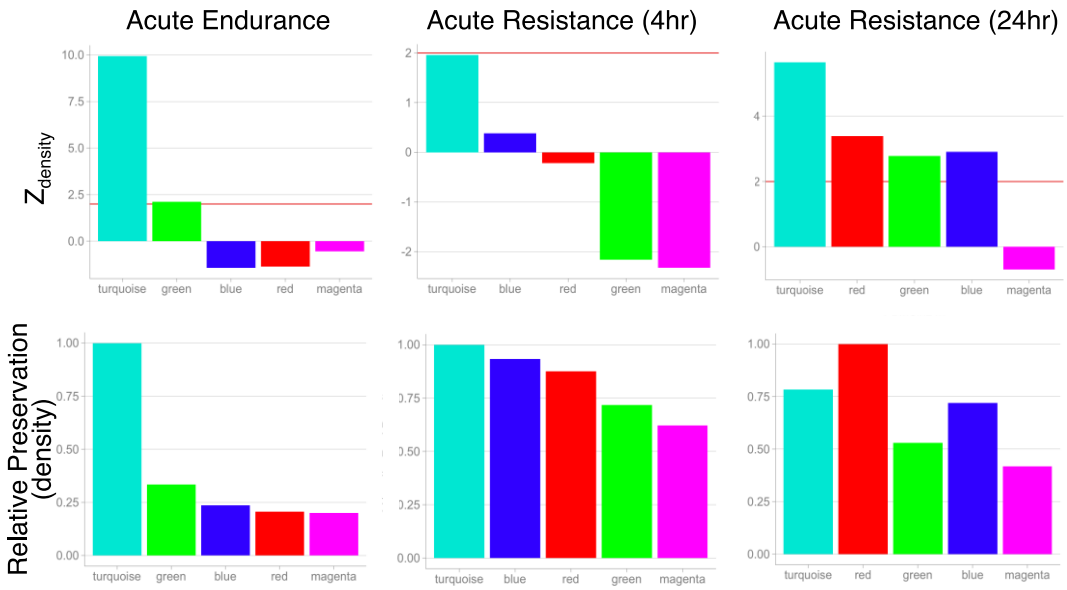


Figure 42.

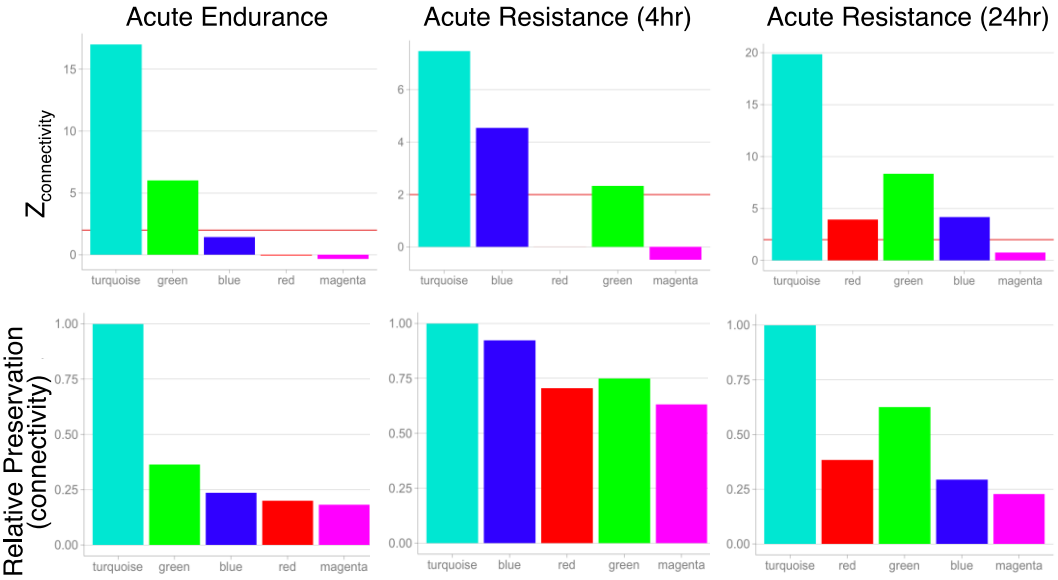


Figure 43.

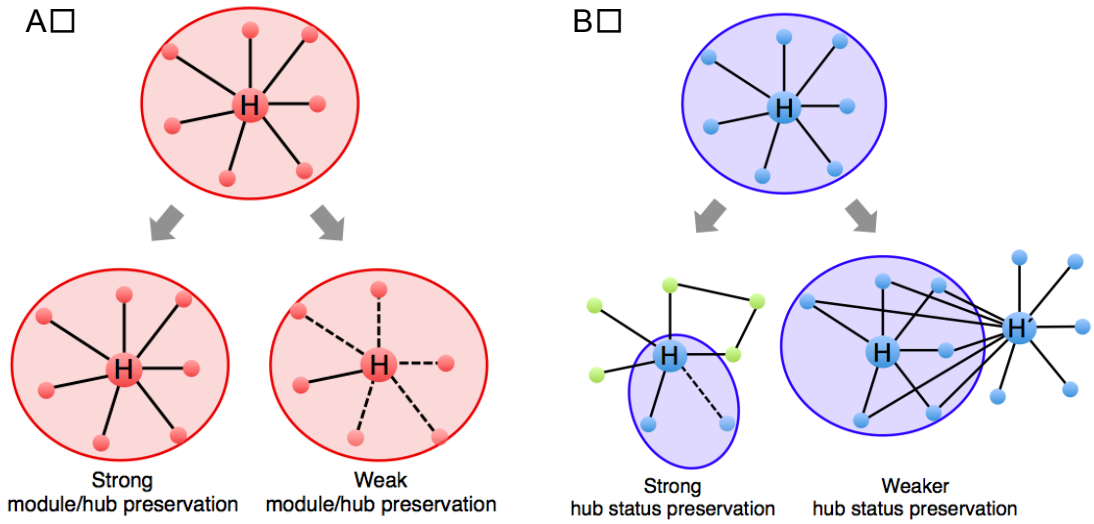


Figure 44.

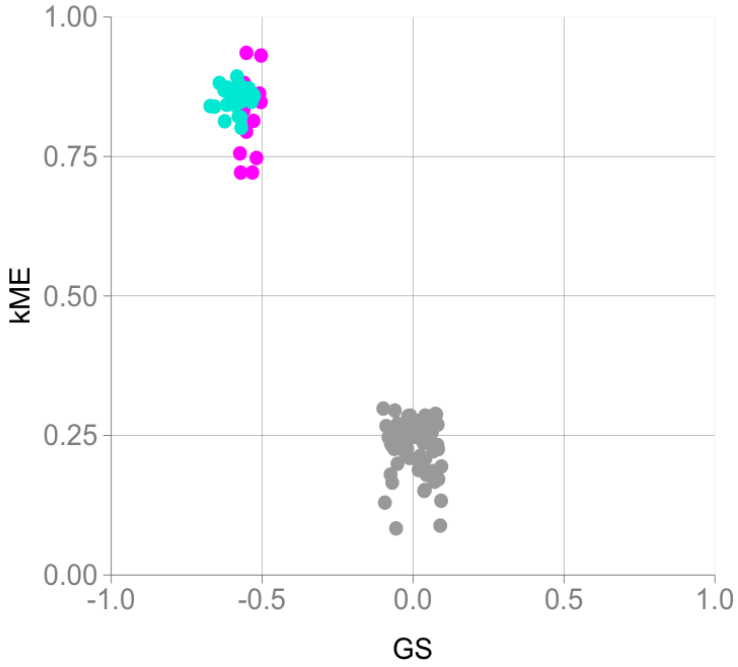


Figure 45.

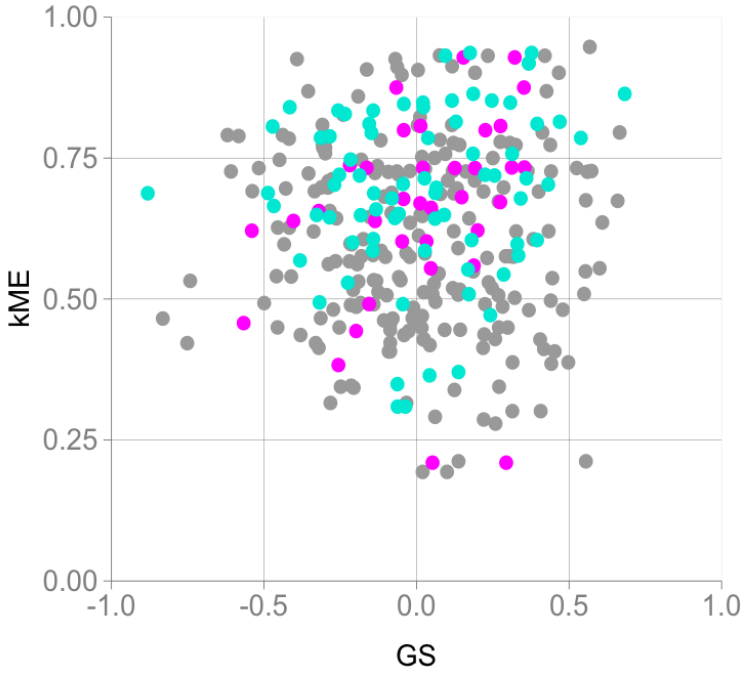


Figure 46.

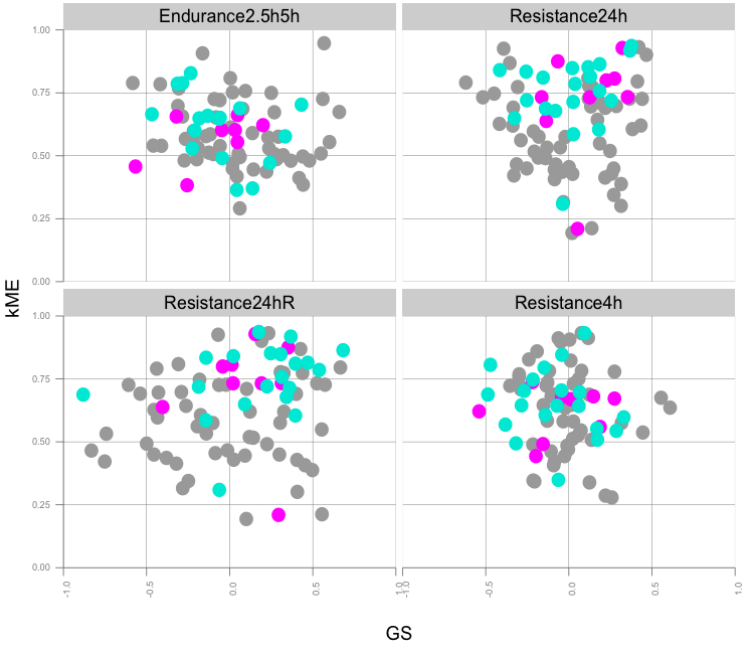


Figure 47.

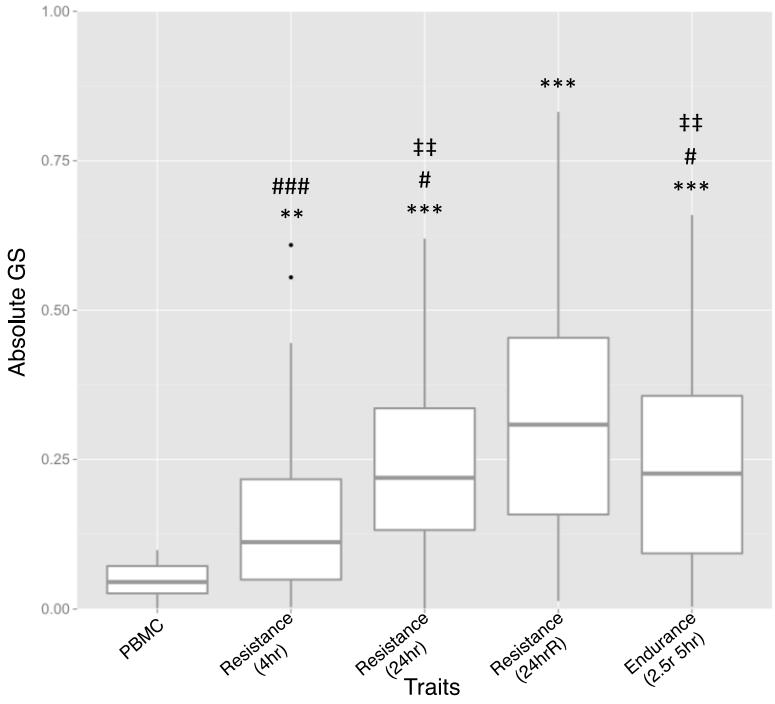


Figure 48.

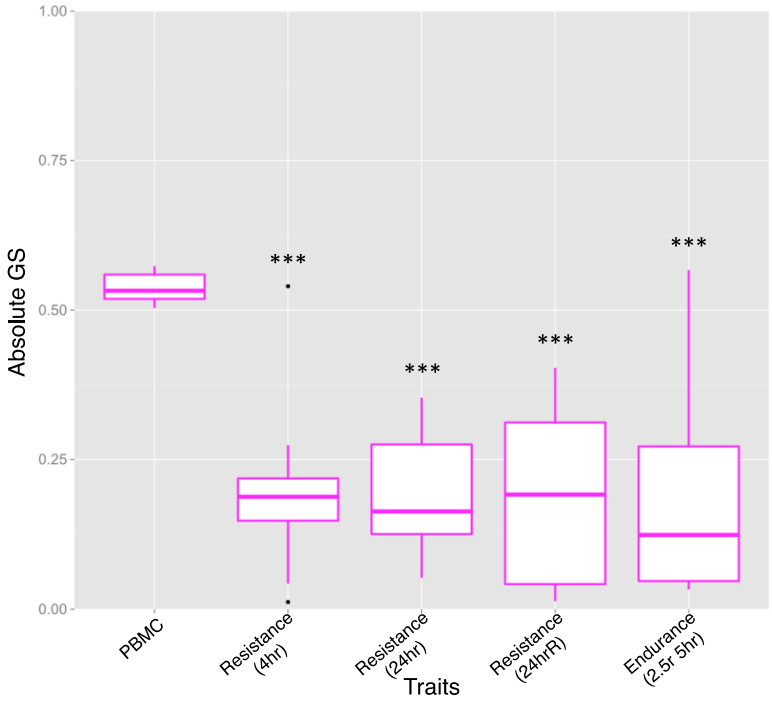


Figure 49.

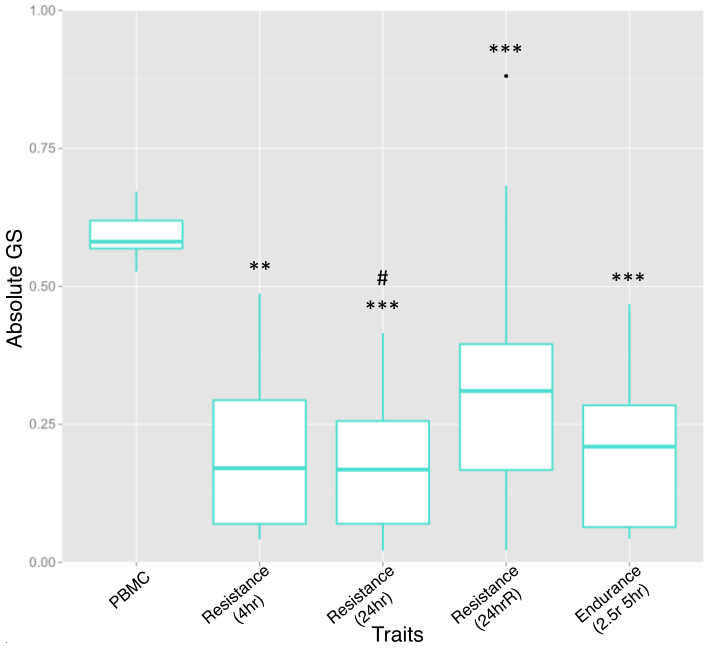


Figure 50.

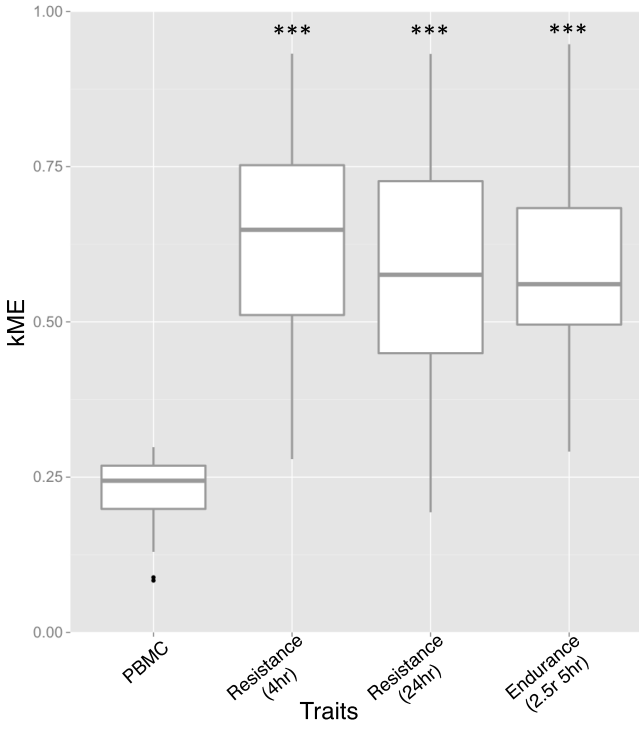


Figure 51.

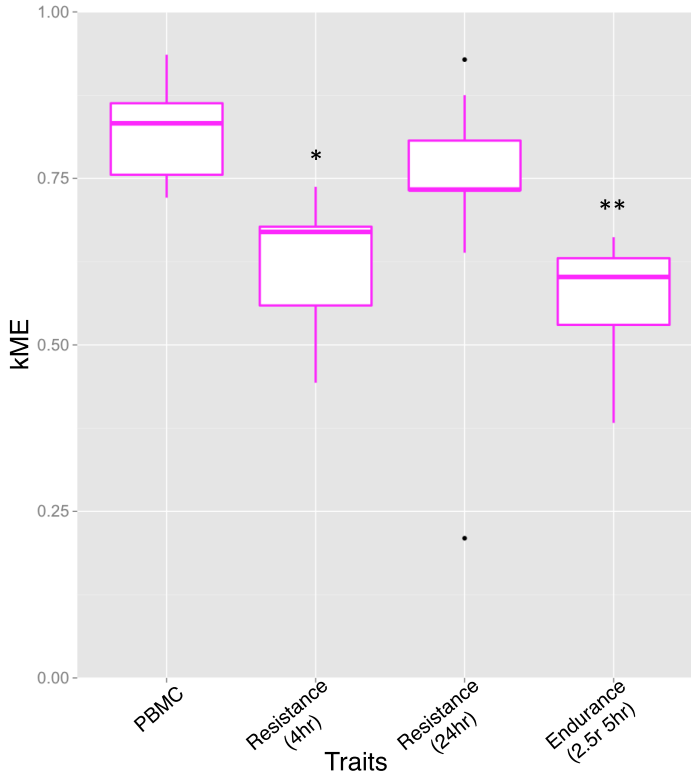


Figure 52.

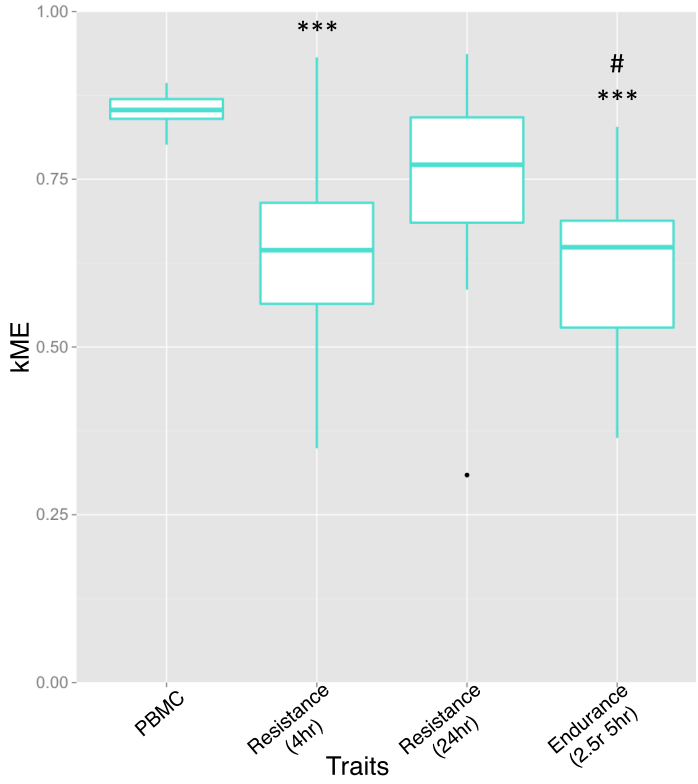


Figure 53.

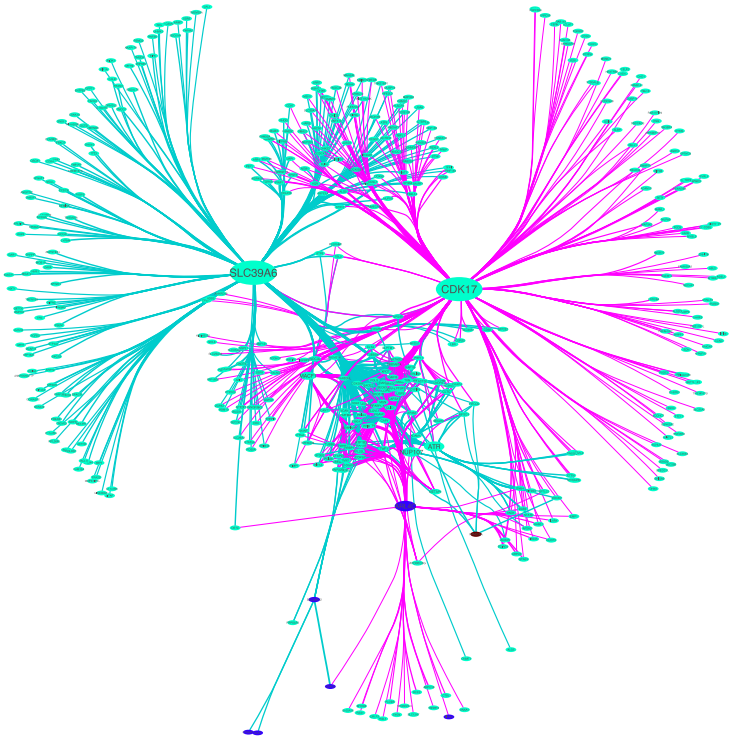


Figure 54.

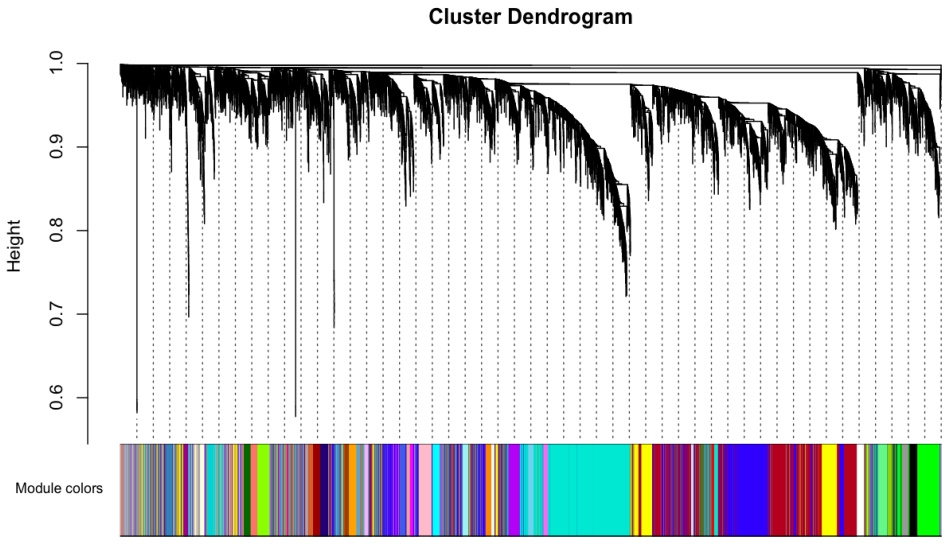


Figure 55.

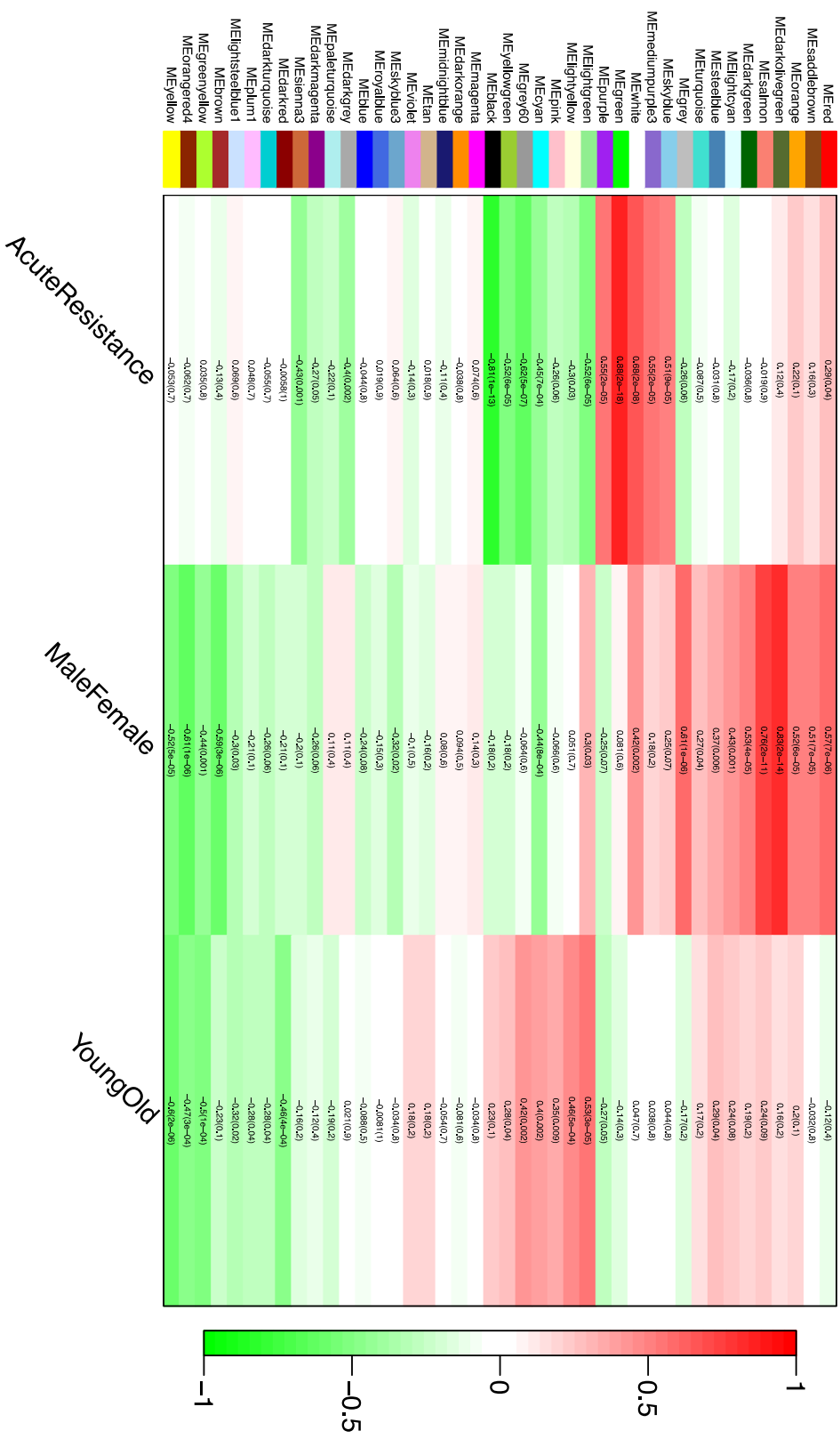

Module−trait relationships
MEred
MEsaddlebrown
MEorange
MEdarkolivegreen
MEsalmon
MEdarkgreen
MElightcyan
MEsteelblue
MEturquoise
MEgrey
MEskyblue
MEmediumpurple3
MEwhite
MEgreen
MEpurple
MElightgreen
MElightyellow
MEpink
MEcyan
MEgrey60
MEyellowgreen
MEblack
MEmagenta
MEdarkorange
MEmidnightblue
MEtan
MEviolet
MEskyblue3
MEroyalblue
MEblue
MEdarkgrey
MEpaleturquoise
MEdarkmagenta
MEsienna3
MEdarkred
MEdarkturquoise
MEplum1
MElightsteelblue1
MEbrown
MEgreenyellow
MEorangered4
MEyellow
AcuteResistance
MaleFemale
YoungOld
1
0.5
0
−0.5
−1

Figure 56.

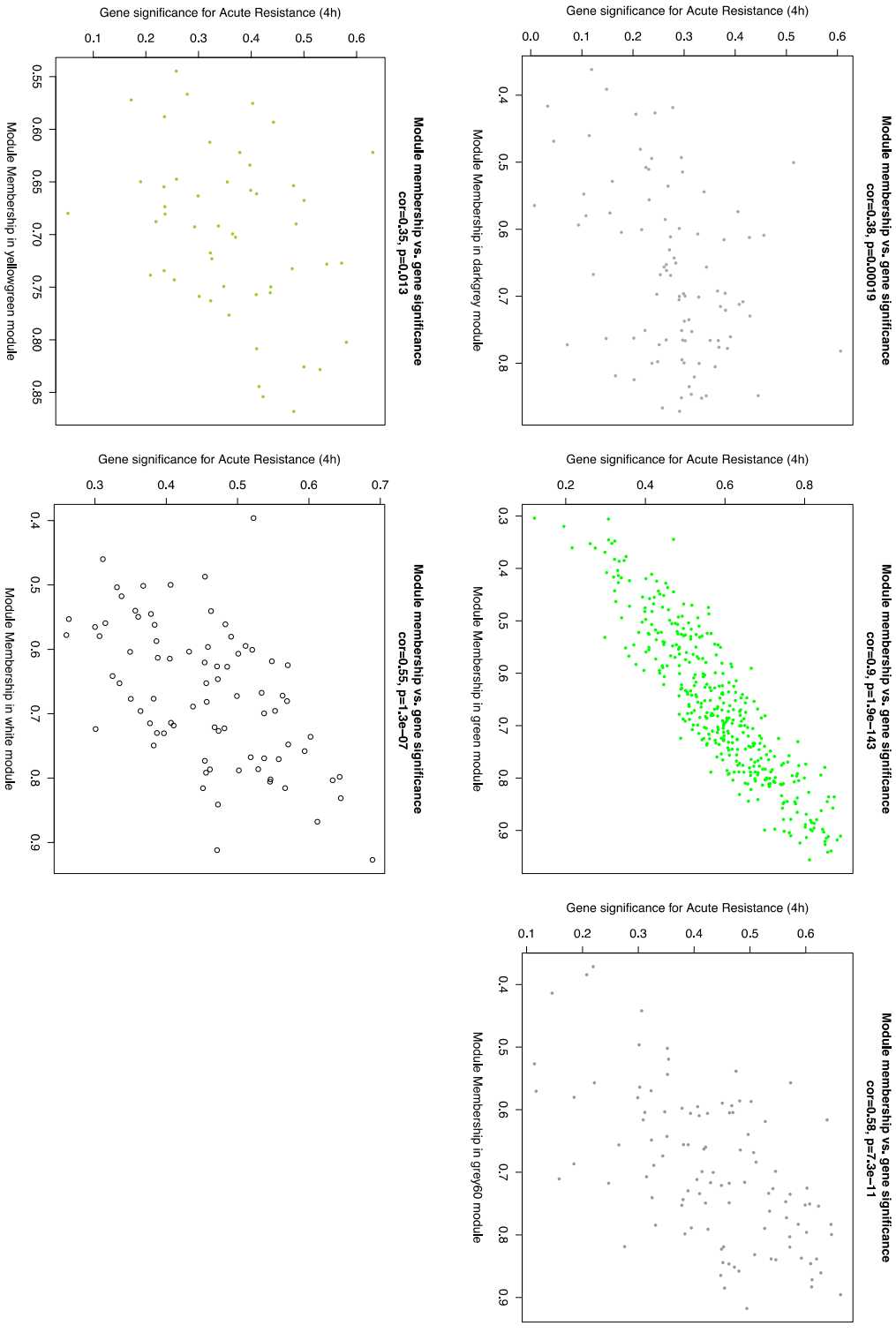

Gene significance for Acute Resistance (4h)
Module membership vs. gene significance
cor=0.38, p=0.00019
Module Membership in darkgrey module
Gene significance for Acute Resistance (4h)
Module membership vs. gene significance
cor=0.35, p=0.013
Module Membership in yellowgreen module
Gene significance for Acute Resistance (4h)
Module membership vs. gene significance
cor=0.9, p=1.9e−143
Module Membership in green module
Gene significance for Acute Resistance (4h)
Module membership vs. gene significance
cor=0.55, p=1.3e−07
Module Membership in white module
Gene significance for Acute Resistance (4h)
Module membership vs. gene significance
cor=0.58, p=7.3e−11
Module Membership in grey60 module

Figure 57.

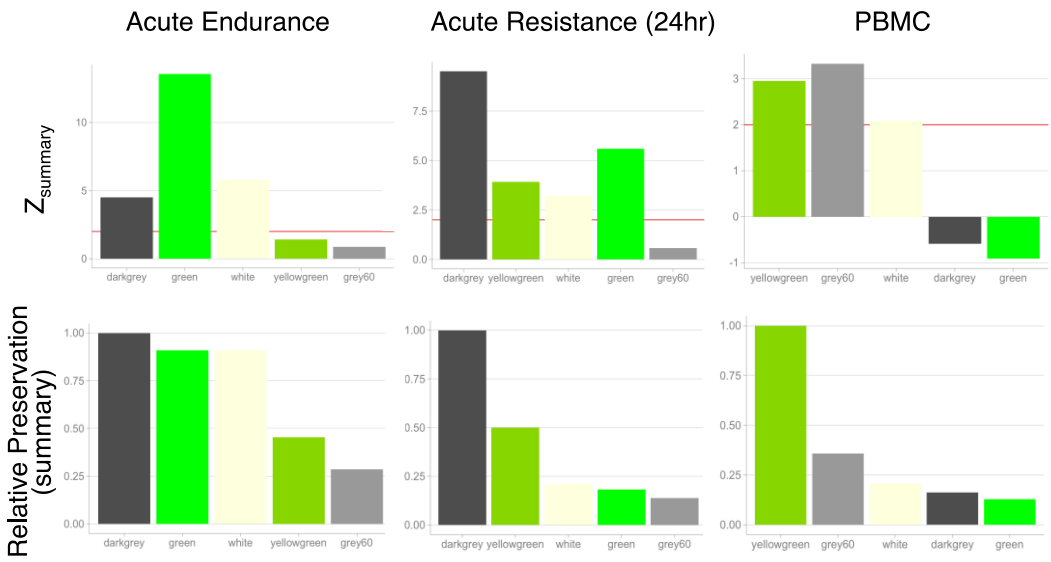


Figure 58.

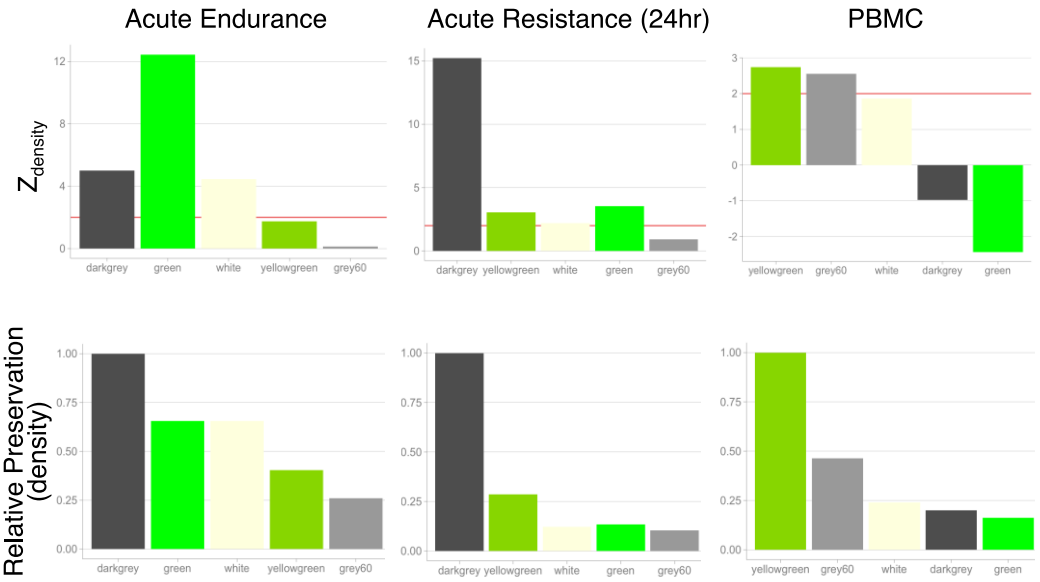


Figure 59.

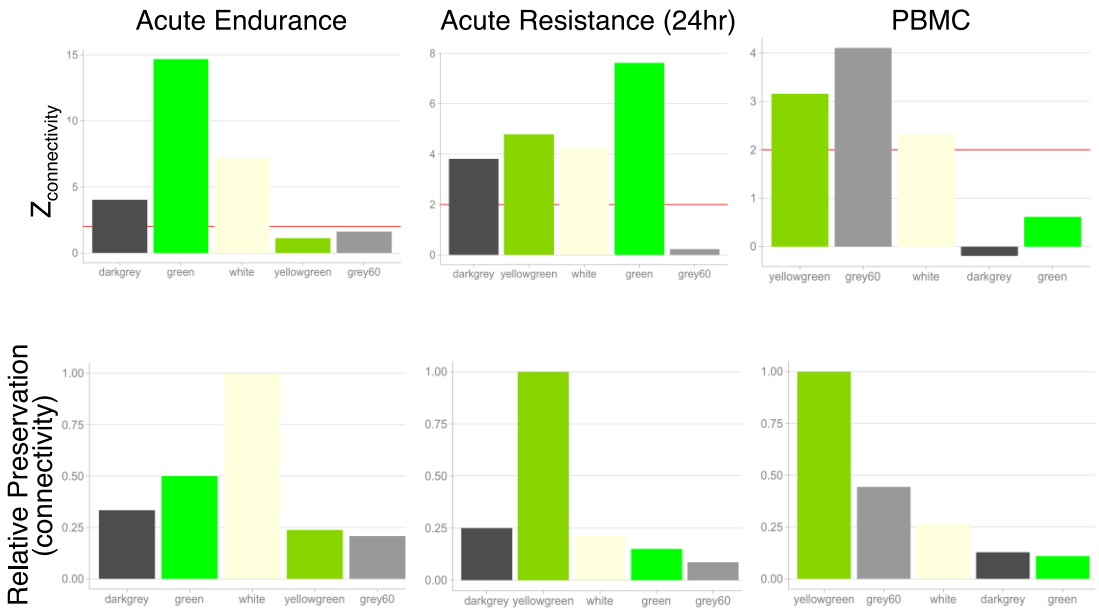

Acute Endurance
Acute Resistance (24hr)
PBMC
Z connectivity
Relative Preservation (connectivity)
darkgrey
green
white
yellowgreen
grey60

Figure 60.

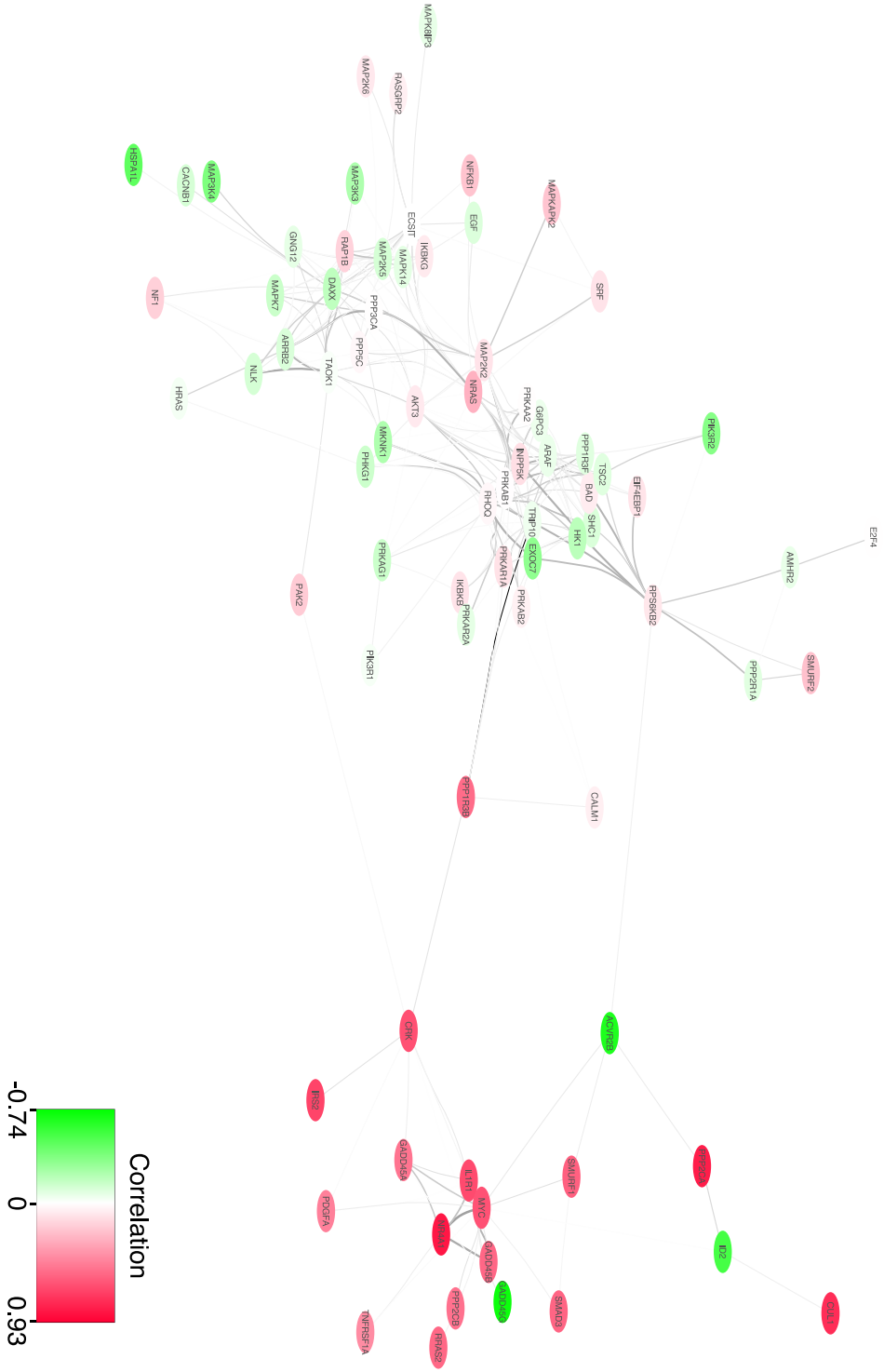


Figure 61.

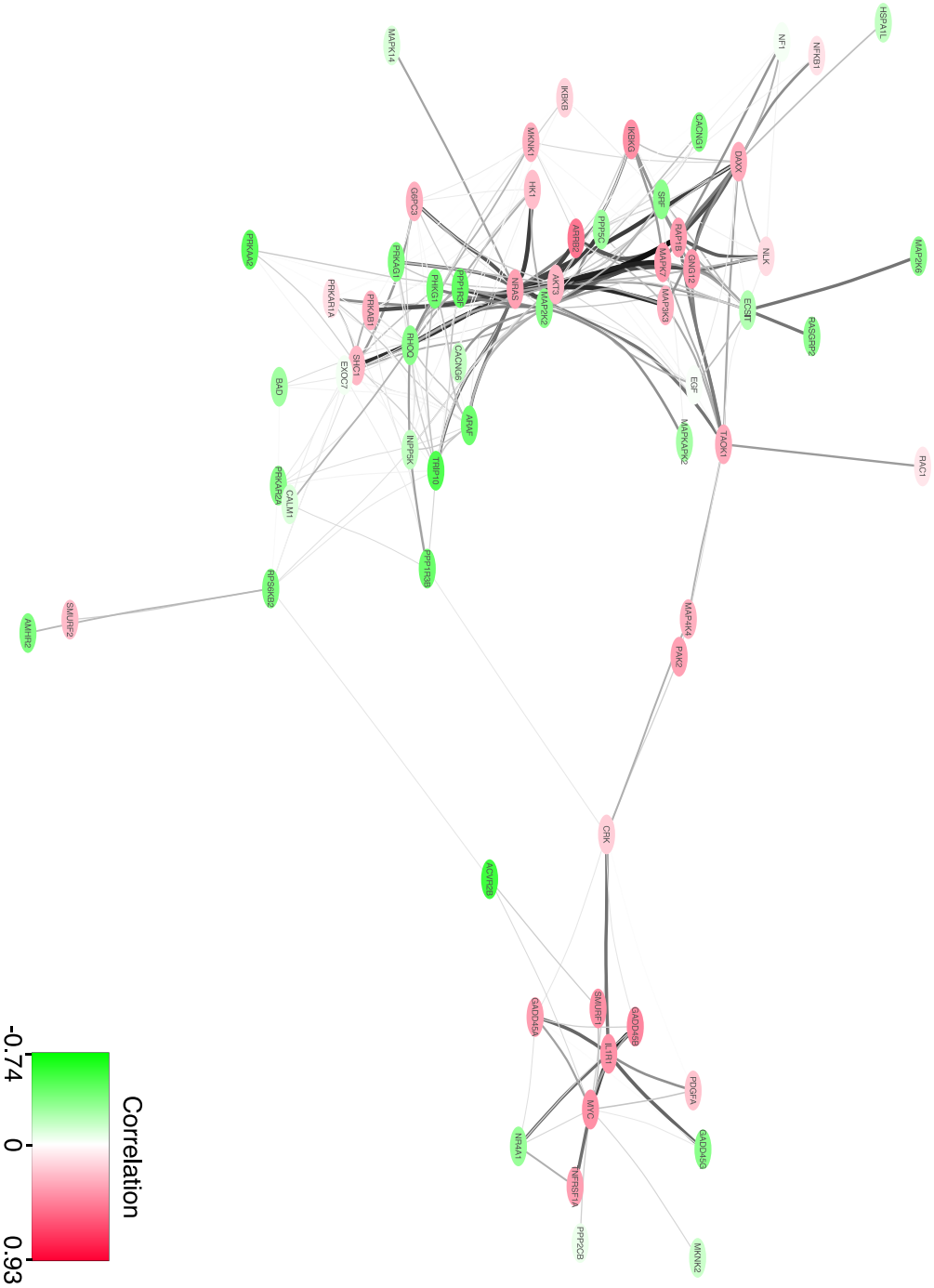

HSPA1L
NF1
NFKB1
MAPK14
IKBKB
MKNK1
IKBKG
CACNG1
DAXX
HK1
SRF
GSTCD
PPP3CC
ARRB2
RAPGEF3
GNG12
NLK
MAP2K6
PRKAB2
PRKAG1
PHKG1
PPP1R8
NRAS
AKT3
MAPK7
MAP3K3
ECSIT
RASGRP2
PRKAR1A
PRKAB1
MAP2R2
RHOQ
SHC1
EXOC7
CACNG6
EGF
BAD
RAF1
MAPKAPK2
TACK1
RAC1
INPP5K
TRIP10
PRKAR2A
CALM1
PPP1R3B
RPS6KB2
SMURF2
AMHR2
MAP4K4
PAK2
CRK
ICYR3
SMURF1
GADD45A
GADD45B
IL1R1
PDGFA
MYC
GADD45G
NR4A1
TNFRSF1A
PPP2CB
MKNK2
-0.74
0
0.93
Correlation

Figure 62.

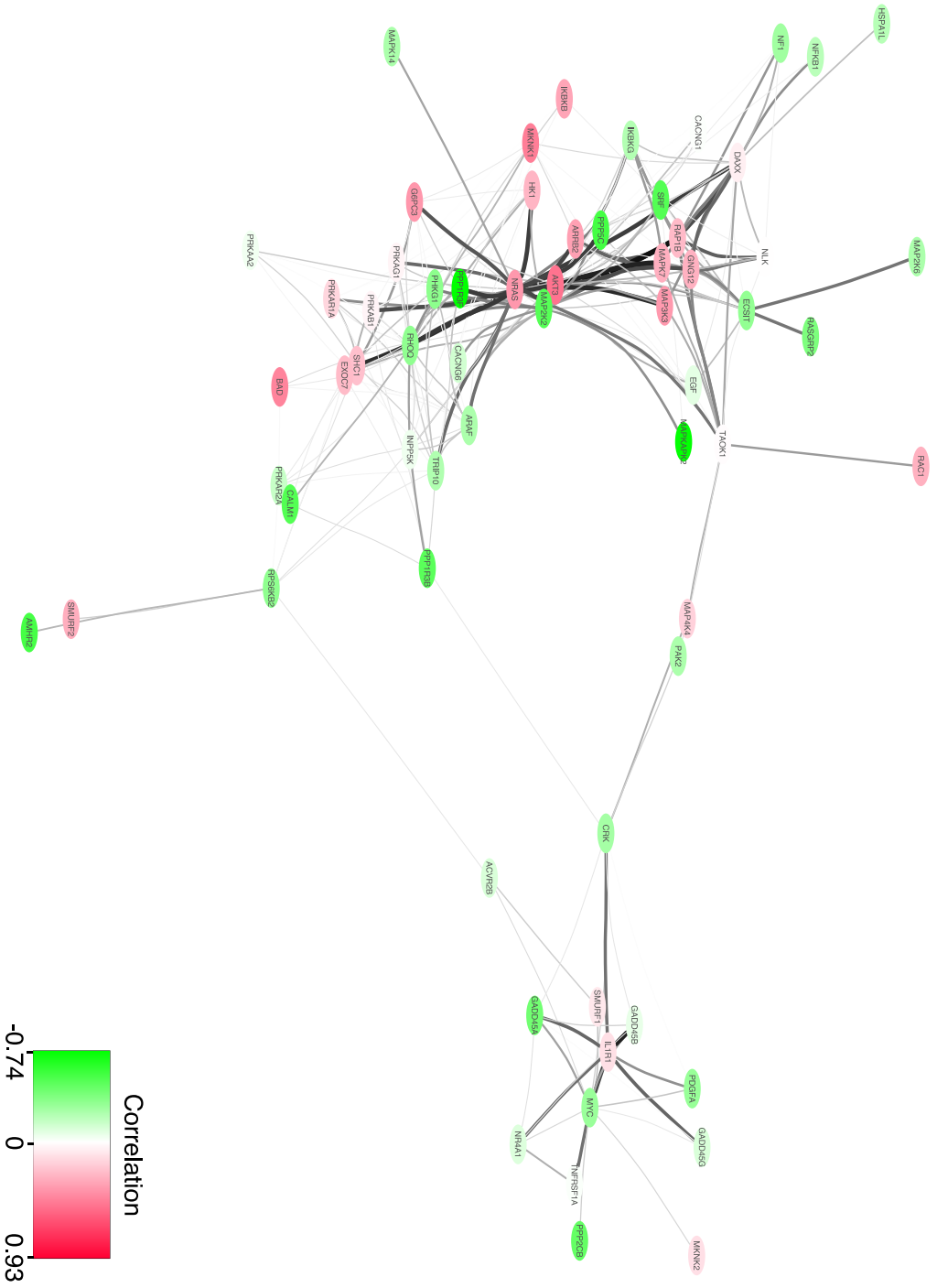

HSPA1L
NF1
NFKB1
MAPK14
IKBKB
MKNK1
IKBKG
CACNG1
DAXX
HK1
SRF
GRPC3
PPP2CA
ARPP21
RAPGEF1
NAPK7
GNG12
NLK
MAP2K6
PRKAA2
PRKAG1
PHKG1
KRAS
AKT3
MAP2K2
MAP2K3
ECSIT
PRKAR1A
PRKAB1
RASGRP2
SHC1
EXOC7
PHKG2
CACNG6
BAD
EGF
ARAF
TAOK1
INPP5K
TRIP10
RAC1
PRKAR2A
CALM1
PPP1R3B
RPS6KB2
SMURF2
MAP4K4
PAK2
CRK
ACVR2B
GADD45A
SMURF1
GADD45B
IL1R1
PDGFA
MYC
GADD45G
NR4A1
TNFRSF1A
PPP2CB
MKNK2
-0.74
0
0.93
Correlation

Figure 63.

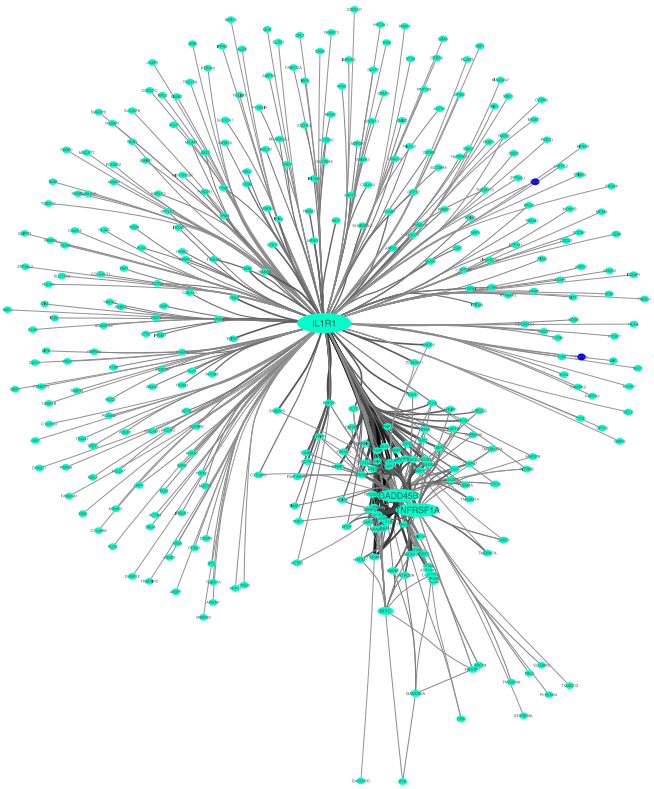


Figure 64.

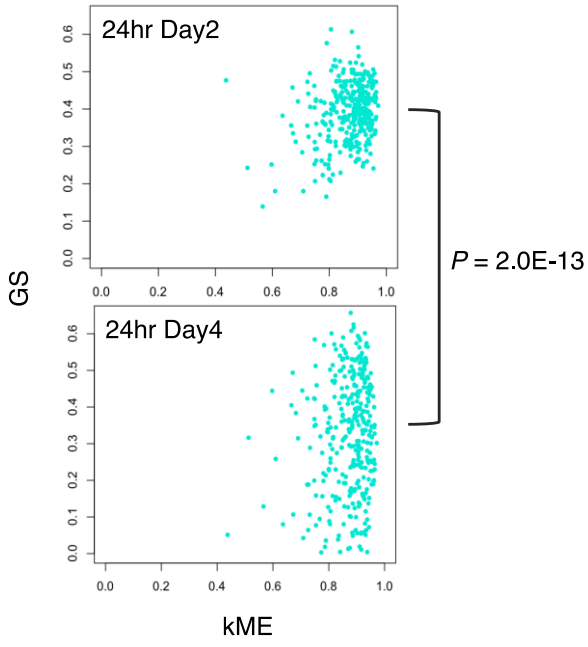


Figure 65.

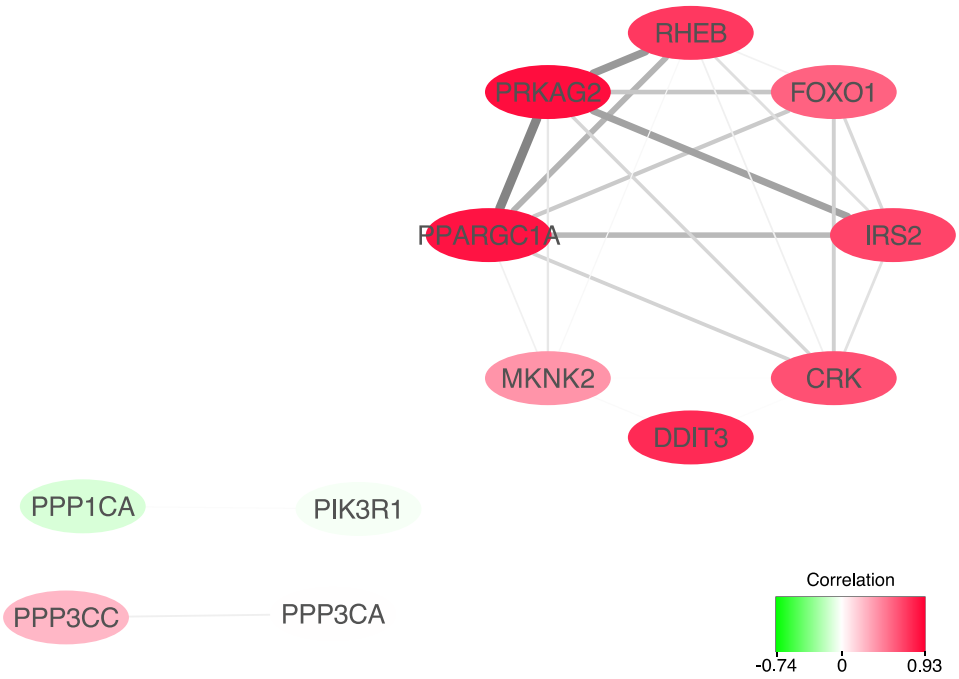


Figure 66.

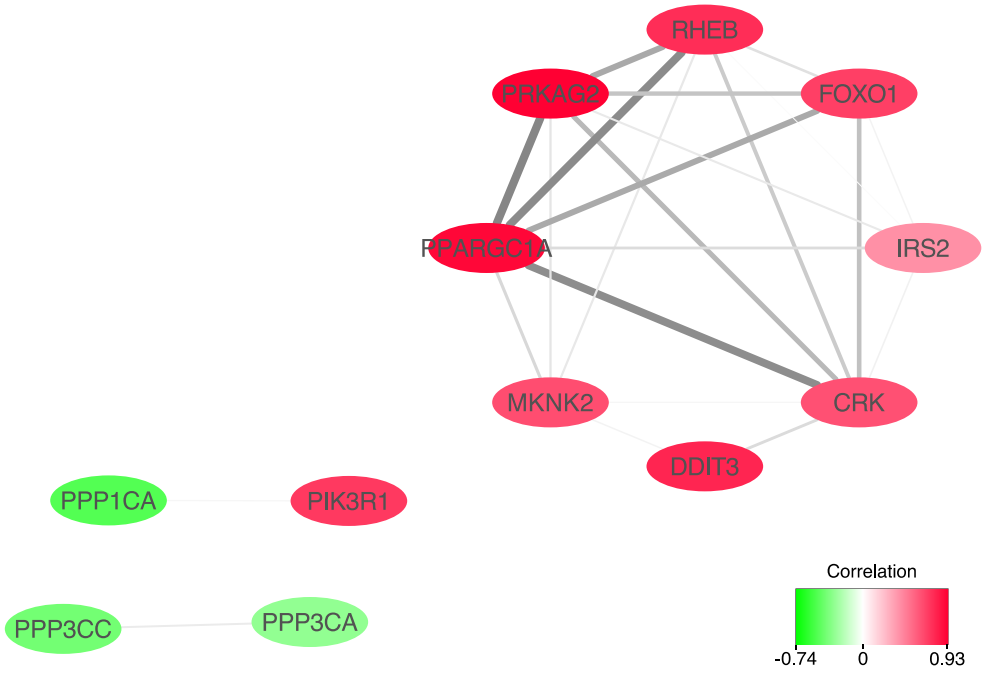


Figure 67.

Table 1. Pathway enrichment in the differentially regulated genes at transcription level

Pathway Enriched in Differentially Expressed Transcription	*P* value (corrected)*
Amino acid synthesis and interconversion (transamination)	0.0024
Alanine, aspartate and glutamate metabolism	0.0039
Fc epsilon receptor (FCERI) signaling	0.032
MAPK signaling pathway	0.060
Innate Immune System	0.15

* *P* values were corrected by Benjamini Hochberg

Table 2. Pathway enrichment in the differentially regulated genes at translation level.

Pathway Enriched in Differentially Regulated Translation	*P* value (corrected)*
Regulation of innate immune responses to cytosolic DNA	0.019
Generic Transcription Pathway	0.023
Alanine, aspartate and glutamate metabolism	0.024
Ion channel transport	0.024
ER-Phagosome pathway	0.040
Glucose metabolism	0.040
Innate Immune System	0.034

* *P* values were corrected by Benjamini Hochberg

Table 3. Gene ontology enrichment in the known TOP motif genes expressed in the current study (> 125 RPM).

Molecular Facntion/Biological Pathway	Enrichment Score
structural constituent of ribosome	68.66
translational elongation	5.71
translational initiation	2.93
ribosome biogenesis	1.48

Table 4. Identified uORFs

ID	mORF gene name	length (aa)	Blast hit	Blast identity
NM_133728	asparagine synthetase domain containing 1 (Asnsd1)	99	asparagine synthetase domain containing 1 [Dasypus novemcinctus]	65%
NM_178602	polymerase (RNA) II (DNA directed) polypeptide M (Polr2m)	84	DNA-directed RNA polymerase II subunit GRINL1A, isoforms 4/5 [homo sapiens]	82%
NM_009870	cyclin-dependent kinase 4 (Cdk4)	50	cyclin-dependent kinase 4 [Elephantulus edwardii]	50%
NM_021473	Alcohol dehydrogenase (Akr1a1)	101	alcohol dehydrogenase [Gluconacetobacter diazotrophicus PAl 5]	40%
NM_008709	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian) (Mycn)	75	PREDICTED: N-myc proto-oncogene protein [Pantholops hodgsonii]	74%

Table 5. Identified Alternative Translation

RefSeq ID	Symbol	Translation	Gene description
NM_001136069	Ldha	Truncated	lactate dehydrogenase A
NM_009415	Tpi1	Truncated	triosephosphate isomerase 1
NM_001163664	Tnnt3	uORF	troponin T3, skeletal, fast
NM_001190451	Dcn	uORF	decorin
NM_007504	Atp2a1	uORF	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1
NM_009813	Casq1	uORF	calsequestrin 1
NM_021285	Myl1	uORF	myosin, light polypeptide 1
NM_198415	Ckmt2	uORF	creatine kinase, mitochondrial 2
NM_007450	Slc25a4	uORF	solute carrier family 25 member 4
NM_025352	Uqcrq	uORF	ubiquinol-cytochrome c reductase, complex III subunit VII

Table 6. Correlation of ribosome densities (residues vs. codons)

Basal	Correlation (Rep1)	Correlation (Rep2)
P	0.8213101	0.8020482
A	0.8402522	0.8579043
1	0.7148192	0.7731411
2	0.7476778	0.7316819
3	0.7427893	0.6608658
4	0.6446547	0.6474076
5	0.3690774	0.3753343

LPS30min	Correlation (Rep1)	Correlation (Rep2)
P	0.7500973	0.7636443
A	0.8215848	0.8107531
1	0.7223672	0.7160933
2	0.7768192	0.7246977
3	0.4714489	0.5407553
4	0.8036243	0.8164233
5	0.4160705	0.41699

Table 7. Module enrichment analysis

Module	Pathway	P-values (corrected)
blue	Lysosome	1.00E-05
green	Antigen processing and presenting	0.02
magenta	NK cell mediated cytotoxicity	1.90E-04
magenta	T cell receptor signaling	6.50E-04
red	Lysosome	2.80E-04
red	Phagosome	0.0023
turquoise	Nucleotide excision repair	0.0027
turquoise	Protein export	0.0041
turquoise	Proteasome	0.0067
turquoise	Ubiquitin mediated proteolysis	0.0071
turquoise	Pyrimidine metabolism	0.011
turquoise	Basal transcription factor	0.013
turquoise	RNA transport	0.029

Table 8. Module Preservation (PBMC)

	turquoise	green	blue	red	magenta
Acte Resistance 4h	**	*	*	–	–
Acute Resistance 24h	***	**	*	*	–
Acute Endurance	***	**	–	–	–

*** strongly preserved (Z > 10)

** modelately preserved (6 < Z < 10)

* weakly preserved (2 < Z < 6)

Table 9. Hub gene network analysis

Connected hub gene module in PBMC	Pathway	P-values (corrected)
magenta	Protein processing in ER	6.30E-03
magenta	Endocytosis	1.20E-02
turquoise	Protein processing in ER	< 1.0E-5
turquoise	Protein export	0.0045
turquoise	RNA transport	0.0089
turquoise	Spliceosome	0.013
shared in magenta and turquoise	Protein processing in ER	< 1.0E-5
shared in magenta and turquoise	Endocytosis	0.046

Table 10. Module enrichment analysis

Module	Pathway	P-values (corrected)
darkgrey	ECM-receptor interaction	< 1.0E-05
darkgrey	Focal adhesion	< 1.0E-05
darkgrey	Protein digestion and absorption	< 1.0E-05
green	TGF-beta signaling	0.026
green	Insulin signaling	0.019
green	MAPK signaling	0.021
grey60	regulation of actin cytoskeleton	0.046
white	Neurotorophin signaling	0.030
yellowgreen	Purine metabolism	0.026

Table 11. Module Preservation (Acute Resistance 4hr)

	green	darkgrey	yellowgreen	white	grey60
Acute Resistance 24h	**	***	*	*	–
Acute Endurance	***	*	–	**	–
PBMC	–	–	*	*	*

*** strongly preserved (Z > 10)

** modelately preserved (6 < Z < 10)

* weakly preserved (2 < Z < 6)

Table 12. Pathway analysis

Pathway	P-values (corrected)
Phagosome	2.50E-04
MAPK signaling	4.50E-04